VISIT…

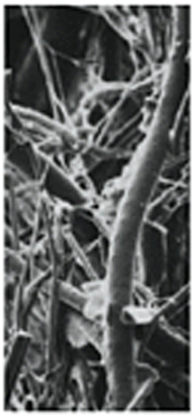
LANZAROTE
Caliente.COM

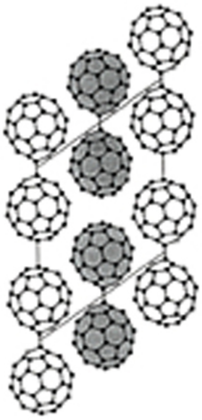
ELSEVIER

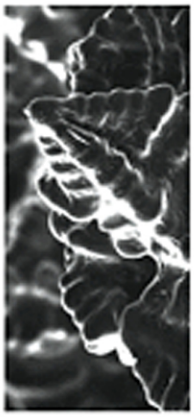
CONCISE ENCYCLOPEDIA OF
THE STRUCTURE OF MATERIALS

Editor

J.W. MARTIN







CONCISE ENCYCLOPEDIA OF

THE STRUCTURE OF MATERIALS

This Page Intentionally Left Blank

CONCISE ENCYCLOPEDIA OF
THE STRUCTURE OF MATERIALS

Editor

J.W. MARTIN
University of Oxford, UK

ELSEVIER

Amsterdam—Boston—Heidelberg—London—New York—Oxford
Paris—San Diego—San Francisco—Singapore—Sydney—Tokyo

Elsevier
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK
Radarweg 29, PO Box 211, 1000 AE Amsterdam, The Netherlands

First edition 2007

Copyright © 2007 Elsevier Ltd. All rights reserved

No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means electronic, mechanical, photocopying, recording or otherwise without the prior written permission of the publisher

Permissions may be sought directly from Elsevier's Science & Technology Rights Department in Oxford, UK: phone (+44) (0) 1865 843830; fax (+44) (0) 1865 853333; email: permissions@elsevier.com. Alternatively you can submit your request online by visiting the Elsevier web site at http://elsevier.com/locate/permissions, and selecting *Obtaining permission to use Elsevier material*

Notice
No responsibility is assumed by the publisher for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions or ideas contained in the material herein. Because of rapid advances in the medical sciences, in particular, independent verification of diagnoses and drug dosages should be made

ISBN-13: 978-0-08-045127-5
ISBN-10: 0-08-045127-6

For information on all Elsevier publications
visit our website at books.elsevier.com

Printed and bound in The Netherlands

07 08 09 10 11 10 9 8 7 6 5 4 3 2 1

Working together to grow
libraries in developing countries
www.elsevier.com | www.bookaid.org | www.sabre.org
ELSEVIER BOOK AID International Sabre Foundation

CONTENTS

Editor's Preface vii

Guide to Use of the Encyclopedia ix

Alphabetical List of Articles xi

Articles 1

List of Contributors 467

Subject Index 473

This Page Intentionally Left Blank

EDITOR'S PREFACE

The Concise Encyclopedia of the Structure of Materials draws together in a single volume a range of articles which appear in the *Encyclopedia of Materials: Science and Technology (EMSAT)*. This very substantial work appeared in 2001, the editors-in-chief being Jurgen Buschow, Robert Cahn, Mert Flemings, Bernhard Ilschner, Ed Kramer and Subhash Mahajan; it has been kept up-to-date by its editors-in-chief in the form of periodic updates loaded on to the internet version.

Materials Science and Technology is concerned with the relationship between the properties and structure of materials. In this context 'Structure' may be defined on many scales, from the sub-atomic (in the case of electrical and magnetic properties, for example), the atomic scale (in the case of crystalline materials), the molecular scale (in polymers, for example), and of course also the microscopic and macroscopic scales. The wide range of experimental techniques available for establishing material structure on these various scales are thoroughly surveyed in *The Concise Encyclopedia of Materials Characterization*, two editions of which have appeared under the editorship of Robert Cahn.

In the present volume, eighty articles have been selected from EMSAT which deal with metals, polymers, ceramics and glasses, biomaterials, wood and paper, liquid crystals, as well as some more general features. Contributions to EMSAT which dealt mainly with the 'properties' of materials (be they physical, mechanical, chemical, and so on), and which made only cursory reference to their structure have of necessity been excluded due to the constraints of the available space.

John W. Martin
Editor

This Page Intentionally Left Blank

GUIDE TO USE OF THE ENCYCLOPEDIA

This Concise Encyclopedia is concerned with the relationship between the properties and structure of materials. Information is presented in a series of alphabetically arranged articles which deal concisely with individual topics in a self-contained manner. This guide outlines the main features and organization of the Encyclopedia, and is intended to help the reader to locate the maximum amount of information on a given topic.

Accessibility of material is of vital importance in a reference work of this kind and article titles have therefore been selected not only on the basis of article content but also with the most probable needs of the reader in mind. An alphabetical list of all the articles contained in this Encyclopedia is to be found on p. xi.

Articles are linked by an extensive cross-referencing system. Cross-references to other articles in the Encyclopedia are of two types: in text and end of text. Those in the body of the text are designed to refer the reader to articles that present in greater detail material on the specific topic under discussion at that point. For example:

> The most widely produced steel product is sheet (see *Sheet Steel: Low Carbon*)…

The cross-references listed at the end of an article serve to identify broad background reading and to direct the reader to articles that cover different aspects of the same topic.

The nature of an encyclopedia demands a higher degree of uniformity in terminology and notation than many other scientific works. The widespread use of the SI system of units has determined that such units be used in this Encyclopedia. It has been recognized, however, that in some fields Imperial units are more generally used. Where this is the case, Imperial units are given their SI equivalent quantity unit following in parenthesis. Where possible, the symbols defined in *Quantities, Units, and Symbols*, published by the Royal Society of London, have been used.

All articles in the Encyclopedia include a Bibliography giving sources of further information. Each bibliography consists of general items for further reading and/or references which cover specific aspects of the text. Where appropriate, authors are cited in the text using a name/date system as follows:

> …another recent review (Padlha *et al.* 2003).
>
> …discovered by Zerr *et al.* (1999)….

The contributors' names and the organizations to which they are affiliated appear at the end of all articles. All contributors can be found in the alphabetical List of Contributors.

The most important information source for locating a particular topic in the Encyclopedia is the multilevel Subject Index, which has been made as complete and fully self-consistent as possible.

This Page Intentionally Left Blank

ALPHABETICAL LIST OF ARTICLES

Aluminum Alloys: Alloy, Heat Treatment, and Temper Designation
Aluminum: Alloying
Annealing Textures
Austenite Decomposition: Overall Kinetics during Isothermal, and Continuous Cooling Transformation
Bainite
Binary Oxide Ceramics: Al_2O_3, ZrO_2, Structure and Properties of
Block Copolymer Phase Behavior
Block Copolymers, Structural Characterization of
Bone Mineralization
Calamitic Liquid Crystals
Carbon Mesophase
Carbon Nanotubes
Cast Irons
Ceramic-modified High-temperature Polymers
Chalcogenide Glasses
Chiral Smectic Liquid Crystals
Cholesteric Liquid Crystals: Defects
Clay-based Polymer Nanocomposites
Cobalt Alloys: Alloying and Thermomechanical Processing
Composite Materials, Microstructural Design of
Copper Alloys: Alloy and Temper Designation
Crystal Engineering
Crystal Structures of the Elements
Crystals, Dendritic Solidification of
Cuticle
Cyclic Polymers, Crystallinity of
Cycloaliphatic Polymers
Deformation Processing: Texture Evolution
Deformation Textures
Discotic Liquid Crystals: Overview
Ferrous Alloys: Overview
Fluorine-containing Polymers
Fullerenes
Glass Ceramics
High-temperature Stable Polymers
Hybrid Dendrimer Star-like Polymers
Ice: Structures
Injection Molded Semicrystalline Polymers: Structure Development and Modeling
Intermetallics: Crystal Structures
Intermetallics: Laves Phases
Liquid Crystals: Overview
Liquid Crystalline Polymers: An Introduction
Magnesium: Alloying
Marine Teeth (and Mammal Teeth)
Martensite
Nanoscale Ceramic Composites
Nematic Discotic Liquid Crystals
Nickel Alloys: Nomenclature
Nickel-based Superalloys: Alloying
Oxide Glasses
Oxynitride Glasses
Paper: Porosity
Paper: Structure
Pearlite
Phase Diagrams
Phase Diagrams: Data Compilation
Polymer Crystallization: General Concepts of Theory and Experiments
Polymer Spherulites
Polymer Surfaces: Structure
Polymer-Ceramic Nanocomposites: Polymer Overview
Polymeric Discotic Liquid Crystals
Porous Ceramics including Fibrous Insulation, Structure and Properties of
Portland Cements
Precipitation from Austenite
Sheet Steel: Low Carbon
Shell: Properties
Si_3N_4 Ceramics, Structure and Properties of
Si-AlON Ceramics, Structure and Properties of
Stainless Steels: Cast
Stainless Steels: Duplex
Steels: Classifications
Structure of Polymer Glasses: Short-range Order
Teeth: Structure/Mechanical Design Strategies
Tissue Engineering Scaffolds
Titanium Alloys: Alloy Designation System
Titanium: Alloying
Tool and Die Steels
Wood, Constituents of
Wood: Macroscopic Anatomy
Wood: Ultrastructure

This Page Intentionally Left Blank

A

Aluminum Alloys: Alloy, Heat Treatment, and Temper Designation

The Aluminum Association of the United States uses a four-digit numerical system to identify aluminum alloys. The system used for wrought alloys is slightly different from that used for cast alloys; however, the digit that designates the alloy group is essentially the same. The nomenclature for wrought alloys has been accepted by most countries and is now called the International Alloy Designation System (IADS). The first digit indicates the alloy group and the last two digits identify the aluminum alloy or indicate the aluminum purity. The second digit indicates modifications of the original alloy or impurity limits. Experimental alloys are also identified in accordance with this system, but they are indicated as experimental by the prefix X. The designation system for wrought and cast aluminum alloys is given in Table 1.

Although the first digit for cast alloys is essentially the same as for wrought alloys, the second two digits serve to identify a particular composition. The zero after the decimal point identifies the product as a casting. Other numerals are used to designate ingots. The 3XX.X series is reserved for alloys that contain silicon and/or additions of copper and magnesium;

Table 1
Aluminum alloy designation systems.

Alloy type[a]	Four-digit designation
Wrought alloys	
99.00% (min) aluminum	1XXX
Copper	2XXX
Manganese	3XXX
Silicon	4XXX
Magnesium	5XXX
Magnesium and silicon	6XXX
Zinc	7XXX
Others	8XXX
Casting alloys	
99.00% (min) aluminum	1XX · X
Copper	2XX · X
Silicon with added copper and/or magnesium	3XX · X
Silicon	4XX · X
Magnesium	5XX · X
Zinc	7XX · X
Tin	8XX · X
Others	9XX · X

[a] Designations are based on aluminum content or main alloying elements.

Table 2
Heat treatment and temper nomenclature for aluminum alloys.[a]

Suffix letter (indicates basic treatment or condition)	First suffix digit (indicates secondary treatment)	Second suffix digit (indicates residual hardening)
F, as fabricated		
O, annealed/wrought products only		
H, cold-worked/work-hardened	1, cold-worked only	2, 1/4 hard
	2, cold-worked and partially annealed	4, 1/2 hard
	3, cold-worked and stabilized	6, 3/4 hard
		8, hard
		9, extra hard
W, solution heat-treated		
T, heat-treated/stable	1, partial solution plus natural aging	
	2, annealed cast products only	
	3, solution plus cold work	
	4, solution plus natural aging	
	5, artificially aged only	
	6, solution plus artificial aging	
	7, solution plus stabilizing	
	8, solution plus cold work and artificial aging	
	9, solution plus artificial, aging and cold work	

[a] Added as suffix letters and digits to the alloy number.

the 6XX.X series is unused, and the 8XX.X series is used for alloys containing tin as the major alloying element. Often a letter prefix is used to denote either an impurity level or the presence of a secondary-alloying element. These letters are assigned in alphabetical sequence starting with A but omitting I, O, Q, and X. X is reserved for experimental alloys. For example A201.0 and A357.0 have higher purity than the original 201.0 and 357.0.

The heat-treatment or temper-nomenclature system developed by the Aluminum Association has also been adopted as part of the IADS by most countries. It is used for all forms of wrought and cast aluminum alloys with the exception of ingot. The system is based on the treatments used to develop the various tempers and takes the form of letters added as suffixes to the alloy number. One or more digits following the letter indicate subdivisions of the tempers, when they significantly influence the characteristics of the alloy. Alloys supplied in the as-fabricated or annealed condition are designated with the suffixes F and O, respectively. The letter W designates those supplied in the solution heat-treated condition. Alloys supplied in the strain-hardened condition are designated with the letter H and those in the heat-treated condition with the letter T. Digits following H represent the degree of strain hardening and those following T the type of aging treatment. The number 5 following the first digit for the age-hardening tempers indicate stress relieving. The temper-designation system is given in Table 2.

See also: Aluminum: Alloying

Bibliography

Aluminum Association 1993 *Aluminum Standards and Data*. The Aluminum Association, Washington, DC

Starke E A Jr 1989 Heat-treatable aluminum alloys. In: Vasudevan A K, Doherty R D (eds.) *Aluminum Alloys—Contemporary Research and Applications, Treatise on Materials Science and Technology*. Academic Press, Boston, MA Vol. 31

E. A. Starke, Jr.
University of Virginia, Charlottesville, Virginia, USA

Aluminum: Alloying

Aluminum is light, ductile, has good electrical and thermal conductivity, and can be made strong by alloying. An advantageous chemical property of aluminum is its reactivity with oxygen, which leads to the formation of a dense layer of Al_2O_3 on the surface that shields the base metal from further environmental interaction. However, problems may be encountered if the layer is disturbed, for example by second phases, plastic deformation that fractures the protective layer, or by friction (tribo-chemical reaction).

Aluminum alloys are classified as heat-treatable or non-heat-treatable, depending on whether they precipitation-harden. However, the properties of both classes of alloys depend on their structural characteristics and can be associated with different levels of structure:

(i) The atom: the low density is due to the low atomic mass ($A_{\mathrm{Al}} = 27$). By alloying with magnesium ($A_{\mathrm{Mg}} = 24$) or lithium ($A_{\mathrm{Li}} = 7$), for example, the density can be reduced still further. Other elements, e.g., silicon ($A_{\mathrm{Si}} = 28$), have little effect.

(ii) The phase: the ductility and formability of aluminum is due to the high symmetry, high stacking fault energy, and thermodynamic stability of the face-centered cubic lattice. Other phases of interest include aluminum solid solutions, intermetallic compounds (Al_3Ti), nonmetallic compounds (AlN), quasicrystals, and metallic glasses.

(iii) The microstructure: the high strength of precipitation-hardened alloys is associated with an ultrafine dispersion of particles ($d < 10\,\mathrm{nm}$). Other microstructural features, e.g., grain boundaries, may have a beneficial effect on strength, but may have a detrimental effect on fracture resistance.

1. *Improving Strength*

The strength of pure aluminum limits its commercial usefulness and a major function of alloying is to improve this property. For structural use, the strongest alloy that meets the minimum requirements for other properties such as corrosion resistance, ductility, toughness, etc., is usually selected if it is cost effective. Consequently, composition is often first selected for strength. The non-heat-treatable alloys obtain their strength from solid-solution strengthening, intermetallic phases that form during solidification, from the dislocation substructure produced by cold or hot work in wrought alloys, and from a fine grain structure in cast alloys. Additions of different solute elements lead to varying degrees of strengthening of the aluminum matrix and magnesium seems to be the most potent solid-solution strengthener. Although solid-solution strengthening can increase the yield strength of pure aluminum several fold, the magnitude is usually below that required for most commercial applications. Phases which form during solidification may also contribute to the strength, but most non-heat-treatable alloys are strengthened through the work hardening that occurs during deformation of the alloys. The solute content, grain size, and volume fraction and size of phases that form during casting can have an effect on the strength increase obtained by work hardening. Most alloys in the non-heat-treatable group contain manganese, silicon, iron, and magnesium, either individually or in various combinations.

The heat-treatable alloys obtain their strength from a homogeneous distribution of fine precipitates developed by solutionizing, quenching, and aging treatments. The increase in strength over that of the pure metal or solid-solution alloys is due to the interaction of dislocations with precipitates and transition precipitates formed during aging. The strength normally increases with the volume fraction of second phases located in the matrix. The relationship between strength and the size of the particles depends on whether the particles are sheared or looped and bypassed by dislocations during deformation. When they are sheared strength increases with increasing particle size and when they are looped and/or bypassed strength increases with decreasing particle size. During normal age hardening there is a decrease in strength associated with a transition from shearing to looping and this is referred to as "overaging." The most important alloying elements in age-hardenable aluminum alloys have extensive solid solubility at high temperatures and include copper, lithium, magnesium, and zinc.

2. *Non-heat-treatable Alloys*

Magnesium is added to non-heat-treatable wrought alloys for its solid-solution strengthening effect. In general, solid-solution alloys are more resistant to corrosion than two-phase alloys. Magnesium makes aluminum more anodic and dilute aluminum–magnesium alloys have a high resistance to corrosion, particularly in seawater and alkaline solutions. However, when magnesium exceeds the solid solubility in binary alloys it precipitates at grain boundaries as Al_8Mg_5, which is strongly anodic and promotes intergranular attack.

Aluminum combines readily with the transition metals titanium, vanadium, chromium, manganese, iron, cobalt, nickel, and zirconium to form intermetallic phases with little or no solubility in the aluminum matrix. Intermetallic phases increase strength by enhancing work-hardening during working operations and by refining the grain structure. They increase work-hardening since they are usually incoherent with the matrix, are nondeformable, and must be looped or bypassed by moving dislocations. This increases the dislocation density and blocks dynamic recovery processes. However, intermetallic compounds greater than 1 μm in diameter, called constituent phases in aluminum alloys, are detrimental since they reduce ductility, fracture toughness, and fatigue-crack-initiation and fatigue-crack-propagation resistance. In addition, dense dislocation tangles develop around these coarse particles during deformation and act as nucleation sites for recrystallization if subsequent annealing operations are required. A suitable distribution of fine particles can prevent growth of recrystallized grains formed at large particles. It is customary to control both composition and cooling rates in order to prevent large primary phases from forming. For example, commercial aluminum–manganese alloys most often contain less than 1.25 wt.% manganese, although as much as 1.82 wt.% is soluble in pure aluminum. The 1.25 wt.% limit is imposed because iron, present in most aluminum alloys as an impurity, decreases the solubility of manganese in aluminum. This increases the probability of forming large primary particles of Al_6Mn, which can have a deleterious effect on ductility.

As mentioned, aluminum has a high stacking-fault energy that aids dynamic recovery and the formation of well-defined subgrains during hot working, making them amenable to substructure strengthening. Strength increases as the subgrain size decreases and the size is controlled by the temperature and strain rate of the hot-working operation. Small intermetallic particles less than 0.6 μm in diameter, which form from the addition of the transition metals previously mentioned, aid in the development and stabilization of the substructure. The intermetallic particles may also add a component of dispersion strengthening, although when concentrations are increased to the point where this becomes a viable strengthening mechanism, coarse primary phases form and ductility is reduced. Rapid solidification of particulate together with powder metallurgy consolidation is one method of circumventing this problem.

Silicon is a principal addition to most aluminum casting alloys because of the high fluidity associated with the aluminum–silicon eutectic. The hard silicon particles are the major strengthening constituents of non-heat-treatable casting alloys. Additional improvements in strength can be obtained by minor additions of magnesium or by other alloy additions, such as sodium, that refine the cast microstructure. The latter also minimizes porosity and increases ductility. The exact mechanism associated with these modifications is unclear, but sodium may depress the eutectic temperature and thus increase the nucleation rate of silicon or tie up phosphorus impurities that are associated with coarse silicon particles. Sometimes, however, phosphorus is added deliberately to hypereutectic aluminum–silicon alloys to refine the size of the primary silicon phases. Phosphorus reacts with the aluminum to form small insoluble particles of AlP that serve as nuclei for the precipitation of silicon. Strontium additions have also been shown to refine the cast microstructure of aluminum–silicon alloys and suppress the formation of primary silicon in hypereutectic alloys. This improves both ductility and toughness.

3. *Heat-treatable Alloys*

Heat-treatable alloys contain elements that decrease in solubility with decreasing temperature and in concentrations that exceed their equilibrium solid solubility at room and moderately higher temperatures. A normal heat-treatment cycle includes a solutionizing soak at a high temperature to maximize solubility, followed by rapid cooling or quenching to a low temperature to obtain a solid solution supersaturated with both solute and vacancies. High strength is obtained by aging for convenient times at one or two intermediate temperatures to form a finely dispersed precipitate. This final step must be accomplished not only below the equilibrium solvus temperature, but also below a metastable miscibility gap called the Guinier–Preston (GP) zone solvus line (Fig. 1) in order to have uniform precipitation. As noted in Fig. 1 the temperature of the line increases with solute content. The critical temperature also increases with vacancy concentration.

The supersaturation of vacancies allows diffusion and clustering of solute atoms to occur much faster than expected from equilibrium diffusion coefficients. In precipitation-hardening alloys, three types of interfaces may develop during nucleation and growth of strengthening precipitates: coherent, semicoherent, and incoherent interfaces. No significant difference in atomic positions exists across a coherent interface, although a slight difference in lattice parameter may result in local elastic strains required to maintain coherency. If the coherency strains become excessive, which usually happens as aging proceeds, the coherent interface may be replaced by a periodic array of edge dislocations or structural ledges resulting in a semicoherent interface. The coherent and semicoherent particles may form in a variety of shapes, e.g., spheres, cubes, disks, and needles, depending on the misfit between particles and the matrix and on the interfacial energy. The shape that occurs generally minimizes the total energy associated with the formation of the particle. For example, when the atomic misfit is very small, spherical particles are formed because this shape has the lowest surface area-to-volume ratio. However, when the atomic misfit is large, the particles generally assume the shapes of disks or needles in order to minimize the strain energy. Cubes are an intermediate shape. The third type of interface, the incoherent interface, has a degree of disorder comparable to that of a high-angle grain boundary.

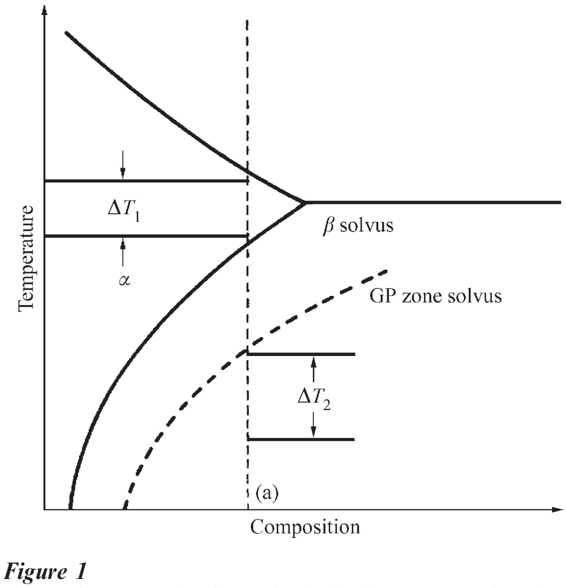


Figure 1
Phase diagram of a hypothetical alloy system showing the β solvus and the GP zone solvus. For composition (a) ΔT_1 is the temperature for solution heat treatment, and ΔT_2 is the temperature range for precipitation heat treatment.

Precipitation normally follows the sequence: supersaturated solid solution → solute clusters → transition precipitate → equilibrium precipitate. Most commercial aluminum alloys contain more than one solute addition to form the strengthening precipitates. Usually, if the atomic size of one solute were larger than the atomic size of the aluminum atom, the other solute addition would be selected to have a smaller atomic size. This minimizes the internal strain associated with clustering of solute atoms, reduces the critical size for precipitation, and results in a more uniform dispersion of strengthening precipitates. Since strength is related to the volume fraction of these particles (in addition to particle size), maximizing those elements that participate in the aging sequence and can be put into solution at the solutionizing temperature, can optimize strength. The solubility of these additions should not be exceeded at the solutionizing temperature since all excesses will form coarse precipitates that do not add significantly to the strength and can adversely affect the fracture toughness and fatigue-crack-initiation and fatigue-crack-propagation resistance.

Often, precipitate-free zones (PFZ) form adjacent to high-angle grain boundaries during the age hardening process. The PFZ may be caused by two different phenomena. In the first, the aging temperature is above the critical temperature determined by the equilibrium concentration of vacancies, but below that determined when excess vacancies are present and vacancies have been lost to the grain boundary prior to aging. Here the PFZ is called a vacancy-depleted PFZ. In the second, normal nucleation and growth of the equilibrium phase occurs at the grain boundary, depleting the adjacent region of solute, and thus producing a solute-depleted PFZ. Both types of PFZs may be minimized by lowering the aging temperature, which decreases the solute supersaturation and thus the driving force for homogeneous decomposition. This also decreases diffusion rates, which decreases the nucleation and growth of the equilibrium precipitates. Consequently, age-hardening alloys are frequently given double aging treatments: a low-temperature age followed by a higher-temperature age. The low-temperature age increases the number of precipitates and minimizes the formation of PFZs and the higher-temperature age accelerates the growth of the precipitates. This allows the desired strength level to be obtained in a minimum amount of time. PFZs can have an adverse effect on the resistance to stress-corrosion-cracking, fatigue-crack-initiation resistance, and ductility and fracture toughness in age-hardenable aluminum alloys.

In addition to the solute elements that are added to form strengthening precipitates, most commercial alloys are designed to have precipitates that control the grain structure. As mentioned, aluminum combines readily with the transition metals, chromium, manganese, zirconium, etc., to form intermetallic phases with little or no solubility in the aluminum matrix. Because of their slow diffusivity in aluminum, these alloying additions form very small precipitates, less than one micron in size, either during solidification or during ingot preheats. This fine distribution of precipitates, called dispersoids, delays or prevents static recrystallization during processing. Because they are strung out in the working direction (a process called mechanical fibering), they also aid in retaining the elongated or ‘‘pancake’’-shaped grains that develop during working. The mean free path of the dispersoids is greater in the working direction than in directions perpendicular to it. Thus, recrystallized grains can grow faster in the working direction. Such elongated grains impart anisotropic behavior to the material. The effectiveness of a particular dispersoid in controlling the grain structure depends on its size, spacing, and coherency.

The first heat-treatable aluminum alloy was based on the aluminum–copper system and resulted from the pioneering work of Alfred Wilm in 1906. After quenching from high temperature and during aging at room temperature, copper atoms cluster into two-dimensional disk-shaped Guinier–Preston (GP) zones on (100) matrix planes. These zones are usually 3–5 nm in diameter. At temperatures of approximately 350 K and higher, three-dimensional zones are formed. Depending on aging time and temperature, they may transform into the transition phase Θ' which is only partially coherent with the matrix. The aging sequence in Al–Cu alloys is usually described as: supersaturated $\alpha \rightarrow \mathrm{GPI} \rightarrow \mathrm{GPII} \rightarrow \Theta' \rightarrow \Theta(\mathrm{Al_2Cu})$. The addition of magnesium to aluminum–copper alloys enhances both the rate and magnitude of natural aging (aging at room temperature). This enhancement probably results from complex interactions between solute elements and vacancies. For high magnesium concentrations the decomposition process may be modified and can be described as: supersaturated $\alpha \rightarrow \mathrm{GP\ zones} \rightarrow \mathrm{S}' \rightarrow \mathrm{S}$, where S′ is the semicoherent, metastable, orthorhombic phase Al_2CuMg and S is the incoherent equilibrium orthorhombic phase. Manganese is usually added to Al–Cu–Mg alloys to form the dispersoid $Al_{20}Cu_2Mn_3$. Aluminum–copper–magnesium alloys are one of the three major heat-treatable aluminum alloy systems.

Aluminum–magnesium binary alloys age-harden following the sequence: supersaturated $\alpha \rightarrow \mathrm{GP\ zones} \rightarrow \beta' \rightarrow \beta(\mathrm{Mg_3Al_2})$. However, most commercial alloys contain additional alloying elements in order to improve such properties as strength and weldability. Medium-strength aluminum alloys, having good weldability and corrosion resistance, can be obtained by adding silicon in balanced amounts to form the quasibinary Al–Mg_2Si, or with an excess of silicon above that needed to form Mg_2Si. These alloys strengthen appreciably during room-temperature aging and also comprise one of the three major heat-treatable aluminum alloy systems. Artificial aging at intermediate temperatures results in the formation of needle-shaped zones. The aging sequence is: supersaturated $\alpha \rightarrow \mathrm{GP\ zones} \rightarrow \beta' \rightarrow \beta(\mathrm{Mg_2Si})$. Small additions of copper are sometimes added to improve mechanical properties and chromium is added to offset the adverse effect that copper usually has on corrosion resistance. Manganese is usually added to Al–Mg–Si alloys to form the dispersoid $Al_{12}Mn_3Si$ for grain structure control.

Alloys based on the Al–Zn–Mg system have a greater response to age-hardening than either the Al–Cu–Mg or the Al–Mg–Si systems. The alloys of commercial interest form spherical GP zones after quenching and during aging from room temperature up to approximately 435K. Extended warm aging of alloys with high zinc:magnesium ratios produces the transition precipitate η', the precursor of the equilibrium $MgZn_2$ η phase. The basal planes of the hexagonal η' precipitate are coherent with the (111) matrix planes, but the interface between the matrix and the *c* direction of the precipitate is incoherent. For low zinc:magnesium ratios the $Al_2Zn_3Mg_3$, T, phase may form. The decomposition process may be described as: supersaturated $\alpha \rightarrow \mathrm{GP\ zones} \rightarrow \eta' \rightarrow \eta$, or supersaturated $\alpha \rightarrow \mathrm{GP\ zones} \rightarrow \mathrm{T}' \rightarrow \mathrm{T}$, where T′ is the semicoherent, metastable, hexagonal phase $Mg_{32}(Al,Zn)_{49}$ and T is the incoherent equilibrium cubic phase $Mg_{32}(Al,Zn)_{49}$.

Copper additions increase considerably the strength of Al–Zn–Mg alloys. Concentrations up to 1 wt.% do not appear to alter the basic precipitation mechanism, and in this range probably add a component of solid-solution strengthening. Low copper-containing alloys are readily weldable, but are not as resistant to stress-corrosion cracking as the higher copper alloys. When in excess of 1 wt.%, copper participates in the precipitation process during aging and decreases the coherency of the precipitate when aged to peak strength. In the quaternary Al–Zn–Mg–Cu system, the phases $MgZn_2$ and AlMgCu form an isomorphous series with aluminum and copper substituting for zinc in $MgZn_2$. The improvement in resistance to stress-corrosion cracking may be related to a reduction in the electrochemical activity of the precipitates as their copper content is increased.

Chromium was originally added to Al–Zn–Mg alloys used for sheet product applications to improve their resistance to stress-corrosion cracking. The chromium-containing dispersoids, $Al_{12}Mg_2Cr$, aid in retaining the directional grain structure developed during processing of wrought products and also prevent excessive growth or formation of recrystallized grains during subsequent heat treatment. However, these incoherent dispersoids increase quench sensitivity. A quenching medium may be forced air, cold or boiling water, or some other liquid. Quench sensitivity describes the tolerance of a particular alloy to the slower quenching media. During slow cooling, some of the solute may precipitate out as coarse particles, which reduces the degree of supersaturation and adversely affects the alloy's fracture toughness and ductility. Normally, quench sensitivity increases with increasing solute content. Chromium-rich dispersoids in high-copper Al–Zn–Mg alloys act as nucleating agents for solute-rich precipitates during quenching. This process eliminates some of the copper and magnesium from participating in the low-temperature aging sequence, and leads to a lower strength than would be predicted from consideration of the alloy chemistry alone. Zirconium may be substituted for chromium, and offers the advantage of forming a coherent dispersoid, Al_3Zr, which minimizes quench sensitivity. Zirconium also seems to be the most effective addition for inhibiting recrystallization.

4. *Special Alloys*

Aluminum-base–lithium ternary alloys, Al–Li–X, are very attractive for applications that require low density and high modulus since each weight percent lithium added to an aluminum-base alloy results in a decrease in density of approximately 3% and an increase in Young's modulus of approximately 6% for lithium additions up to the practical limit of 3 wt.%. The beneficial effects of lithium are realized regardless of whether or not lithium is retained in solid solution, but binary alloys containing in excess of 1% lithium respond to age-hardening due to the precipitation of the metastable phase Al_3Li. The precipitates are ordered and coherent with the aluminum matrix, similar to those in other age-hardenable aluminum systems. Magnesium is a potent solid-solution strengthener when added to aluminum–lithium alloys and also decreases the solubility of lithium in the aluminum matrix and thus enhances the age-hardening response. There are a number of commercial Al–Li–Mg alloys that, due to their low density, are especially attractive for aerospace applications. Copper has also been added to aluminum–lithium alloys with and without magnesium and has a significant strengthening effect. In some alloys a number of strengthening phases coprecipitate, e.g., Al_2Cu, Al_2CuLi, and Al_3Li. These alloys can have very high strengths, sometimes exceeding those obtained in Al–Zn–Mg alloys. As a class, the Al–Li–X alloys appear to have the best overall fatigue-crack-growth resistance of any of the age-hardenable aluminum alloys and are being used in some fatigue-critical structural components of aircraft, e.g., the US Lockheed F-16 fighter aircraft. Most Al–Li–X alloys contain zirconium as the dispersoid-forming element but some also have manganese, in addition to zirconium, to form a dispersoid to minimize strain localization, which is prevalent in this class of alloys.

Calcium additions also lower the density of aluminum and aluminum–calcium alloys are superplastic at high temperatures and medium strain rates. The aluminum-rich end of the aluminum–calcium phase diagram has a eutectic that forms microduplex structures, with the Al_4Ca phase precipitating as rods or lamellae, depending on the solidification rate. The term superplastic is used to describe alloys that may be deformed by over 200% and exhibit large tensile elongation without necking. Superplastic deformation is usually conducted at temperatures above one half of the melting temperature of the alloy, at forming stresses of approximately 4–6 MPa and at strain rates of approximately $10^{-4}\,s^{-1}$. Superplastic alloys are of commercial interest because of the possibility of forming complex shapes using simple manufacturing operations. Many aluminum alloys having eutectic and eutectoid reactions exhibit the phenomenon. Alloys designed especially for superplastic forming are commercially available and include eutectic alloys of Al–Cu–Zn and Al–Cu–Zr. Some complex aluminum systems (e.g., Al–Zn–Mg–Cu and Al–Mg–Si) can be made superplastic by thermomechanically treating the material to achieve a very fine, equiaxed, and stable grain size.

5. *Trace Additions*

Trace elements may often be present in aluminum alloys, either as impurities or by design, and can have both beneficial and adverse effects on the microstructure and properties of the aluminum alloys. Elements such as bismuth, cadmium, and lead have limited solubility in aluminum and may improve machinability. Copper, magnesium, silicon, and zinc may also improve machinability if the precipitates formed are not too hard. Cadmium, in addition to indium, tin, and silver among others, can modify the shape, size, and distribution of second phases and thus affect the characteristics of the alloy. These modifications arise from the influence that trace elements have on: (i) the critical temperature for uniform precipitation; (ii) vacancy concentration and diffusion rates; (iii) interfacial energies; and (iv) precipitate type.

Silicon normally increases fluidity, feedability, and resistance to hot cracking. Iron reduces hot cracking in aluminum–copper and aluminum–zinc alloys but both iron and silicon can form coarse constituent phases that reduce fracture toughness and fatigue-crack-initiation and fatigue-crack-propagation resistance. Consequently, when these properties are of prime importance, iron and silicon contents must be kept to a minimum. Small amounts of titanium, and titanium plus boron, refines the cast structure of aluminum alloys and improves the resistance to hot cracking. However, these additions also reduce fluidity. During casting of some aluminum alloys, especially those containing magnesium, deleterious oxides may be entrapped. This may be prevented by adding trace amounts of beryllium, which forms a thin protective oxide film over the surface of the molten metal and also prevents the loss of magnesium from the melt.

See also: Aluminum Alloys: Alloy, Heat Treatment, and Temper Designation

Bibliography

Blankenship C P Jr, Starke E A Jr, Hornbogen E 1996 Microstructure and properties of aluminum alloys. In: Li J C M (ed.) *Microstructure and Properties of Materials*. World Scientific, Singapore, Vol. 1, pp. 1–49

Humphreys F J, Hatherly M 1995 *Recrystallization and Related Annealing Phenomena*. Pergamon, New York

Lorimer G W 1978 Precipitation in aluminum alloys. In: Russell K C, Aaronson H I (eds.) *Precipitation Processes in Solids*. The Metallurgical Society (AIME), New York, pp. 87–119

Polmear I J 1995 *Light Alloys: Metallurgy of the Light Metals*, 3rd edn. Arnold, London pp. 35–63

Ryum N 1986 Physical metallurgy of heat-treatable alloys. In: Starke E A, Jr Sanders T H Jr (eds.) *Aluminum Alloys—Their Physical and Mechanical Properties*. EMAS, West Midlands, UK, Vol. III, pp. 1511–46

Sanders R E Jr, Baumann S F, Stumpf H C 1986 Non-heat-treatable aluminum alloys. In: Starke E A, Jr Sanders T H Jr (eds.) *Aluminum Alloys—Their Physical and Mechanical Properties*. EMAS, West Midlands, UK, Vol. III, pp. 1441–84

Starke E A Jr 1970 The causes and effects of "denuded" or "precipitate-free" zones at grain boundaries in aluminum base alloys. *J. Met.* **22**, 54–63

Starke E A Jr 1989 Heat-treatable aluminum alloys. In: Vasudevan A K, Doherty R D (eds.) 1989 *Aluminum Alloys—Contemporary Research and Applications, Treatise on Materials Science and Technology*. Academic Press, Boston, MA Vol. 31

Starke E A Jr, Staley J T 1996 Application of modern aluminum alloys to aircraft. *Prog. Aerosp. Sci.* **32**, 131–72

E. A. Starke, Jr.
University of Virginia, Charlottesville, Virginia, USA

Annealing Textures

During annealing of deformed materials three fundamental processes can take place: recovery, recrystallization, and grain growth. These processes may take place consecutively or simultaneously and are differentiated by their driving forces. The crystallographic texture, i.e., the preferred crystallographic orientations, may or may not change during these processes.

(i) Recovery generally does not lead to very significant texture changes and the deformation texture (see *Deformation Textures*) is thus largely maintained during this process. Measurable changes in the texture may be observed if subgrain coarsening occurs during recovery. In this case, for example, some minority texture components can be eliminated.

(ii) During recrystallization new comparatively strain-free nuclei form and grow at the expense of the deformed material until it is consumed. Recrystallization may be static or dynamic. Static recrystallization occurs upon subsequent annealing of a deformed sample, while dynamic recrystallization refers to the occurrence of recrystallization during deformation. Very large changes in texture may take place during recrystallization. An example is shown in Fig. 1. Here the deformation texture is completely replaced by an entirely different texture. In other cases, no texture change is observed during recrystallization. This is, for example, often observed when large volume fractions of second-phase particles are present. Important parameters determining the occurrence, or lack, of texture change during recrystallization are materials parameters (e.g., initial grain size, alloy composition, presence of precipitates) and process parameters (e.g., deformation mode, strain, deformation temperature). Parameters related to the annealing, such as heating rate and temperature, may also affect the texture development, but are generally of less importance.

(iii) During grain growth the recrystallized grains coarsen either by normal grain growth, where the microstructure coarsens in a uniform way, or by abnormal grain growth, where only a few grains in the recrystallized matrix grow very large at the expense of the others, until these large grains impinge upon each other. Significant texture changes may take place during grain growth. Such changes are generally very pronounced after abnormal grain growth, whereas normal grain growth may or may not lead to texture changes. Parameters important for the texture change during grain growth are the presence of solutes and second-phase particles, recrystallization texture, and annealing temperature. If the sizes of the grains become comparable to at least one of the specimen dimensions the annealing atmosphere is also important.

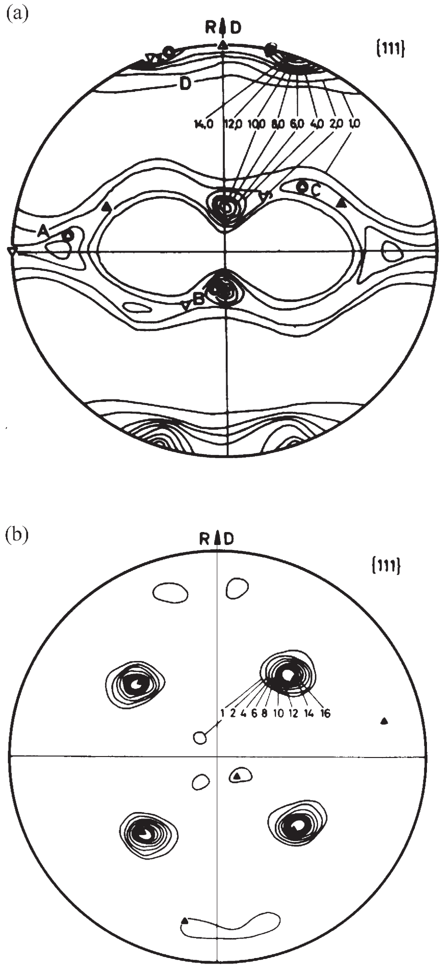


Figure 1
{111} pole figures for pure copper. Contour lines and numbers are used to show the strength of the texture (a) after rolling to 95% reduction in thickness and (b) after recrystallization (after Schmidt and Lücke 1979).

1. Typical Examples of Annealing Textures in Metals and Alloys

The number of measurements of annealing texture is enormous. In this article some typical results for rolled sheets and drawn wires of selected metals are reported. For more information on these and other materials see, for example, Wassermann and Grewen (1962), Barrett and Massalski (1980), and (Wenk 1985).

1.1 Rolled Sheets, f.c.c.

In heavily rolled sheets of f.c.c. metals with medium to high stacking fault energy (SFE), recrystallization typically leads to development of the cube texture (see Figs. 1 and 2) with a {100} plane parallel to the rolling plane and a ⟨100⟩ direction along the rolling direction. The lower the SFE the more cube twin orientations are present in the recrystallization texture. The cube texture is often seen admixed with rolling components (i.e., texture components present in the deformation texture after rolling) and so-called random components (i.e., other texture components fairly randomly distributed in orientation space). The development of the cube recrystallization texture depends significantly on addition of alloying elements. For example, no cube texture is observed in copper when 5 wt.% zinc, 0.5 wt.% cadmium, or 0.0025 at.% phosphorus is added. Technologically the cube texture is very important because it is associated with strong anisotropy in the mechanical (e.g., earing) and magnetic properties.

In sheets of f.c.c. metals or alloys of low SFE the cube texture does not develop during recrystallization. Instead a texture evolves with a main component {236} ⟨385⟩ which is near {113} ⟨211⟩.

Upon further annealing to grain growth, texture sharpening is observed in many f.c.c. metal sheets of both high and low SFE. A complete change of the recrystallization texture may occur during abnormal grain growth. For example, in copper sheets the cube recrystallization texture is replace by new orientations generated by a 30–40° ⟨111⟩ rotation and in silver and brass (30% zinc) the rolling texture component {110} ⟨112⟩ returns during abnormal grain growth.

1.2 Rolled Sheets, b.c.c.

In many b.c.c. metals including iron and several steels, the deformation-type texture is maintained during recrystallization but the concentrations of the various texture components can be significantly changed and depend on alloying as well as annealing parameters. The recrystallization texture typically contains a whole range of components with {111} planes almost parallel to the rolling plane and several components with a ⟨110⟩ direction parallel to the rolling direction. This texture is described as a fairly complete {111} ND fiber texture and a partial ⟨110⟩ fiber texture (ND = normal direction). The {111} ND fiber texture is generally the stronger of the two and along this fiber the {111} ⟨112⟩ and/or {111} ⟨110⟩ components are often most pronounced. An example is shown in Fig. 3. The {111} ND fiber texture is technically very important in order to obtain good, deep drawability.

Grain growth generally leads to strengthening of the {111} fiber texture. A special example is abnormal grain growth in silicon steel sheets. During this process, a strong Goss texture with a {110} plane parallel to the rolling plane and a ⟨100⟩ direction along the

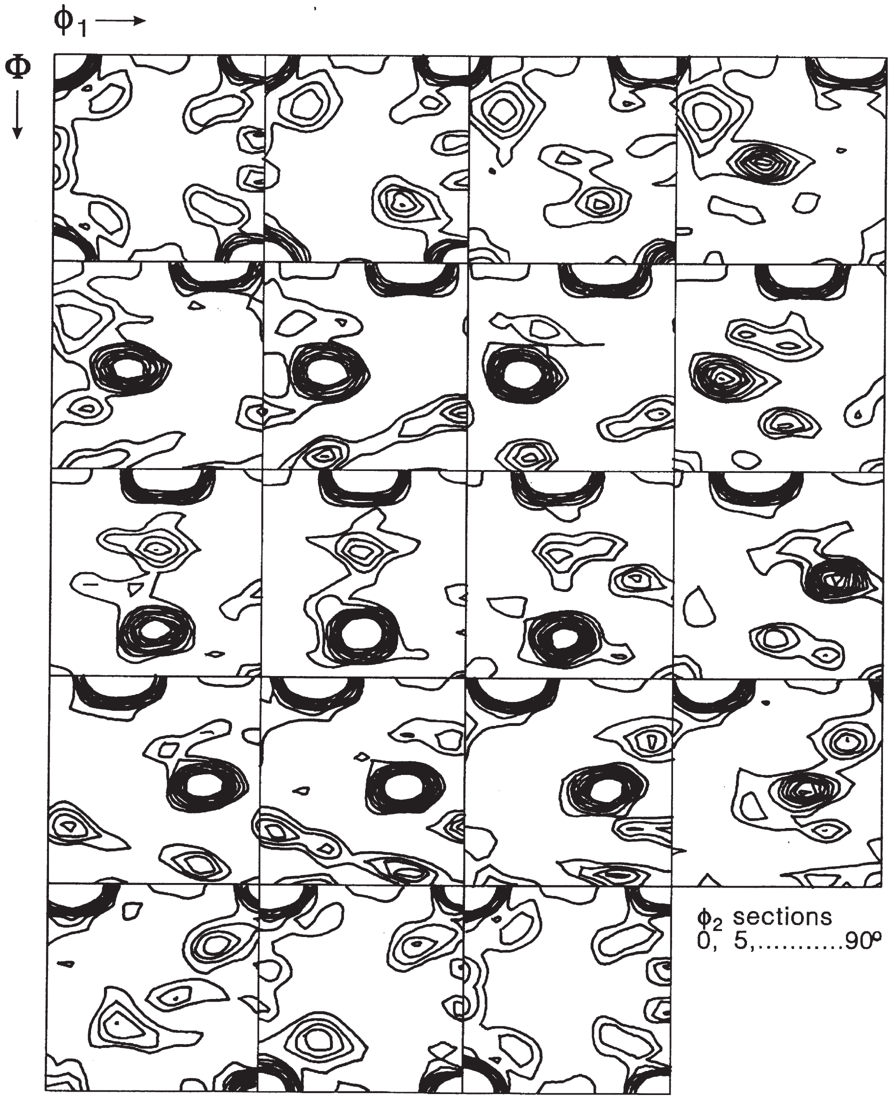


Figure 2
Orientation distribution function for recrystallized pure copper showing a marked cube texture. The contour levels are 1, 2, 3…x random.

rolling direction is developed. Silicon steels are used for transformer cores and, as magnetization is easiest in the ⟨100⟩ directions, great effort is devoted to strengthening such texture components (e.g., the Goss texture) to improve efficiency.

1.3 Rolled Sheets, c.p.h.

The recrystallization textures in hexagonal metal sheets are in general retained rolling textures. There are, however, also examples of texture changes during recrystallization: annealing of titanium in the α-phase gives a $(02\bar{2}5)[2\bar{1}\bar{1}0]$ recrystallization texture which is related to the main rolling component by a 30° [0001] rotation; and zirconium annealed at 400–600 °C develops new recrystallization components with the $[11\bar{2}0]$ direction near the rolling direction.

1.4 Drawn Wires, f.c.c., b.c.c., c.p.h.

In wires of f.c.c., b.c.c., and c.p.h. metals the deformation texture is generally retained during

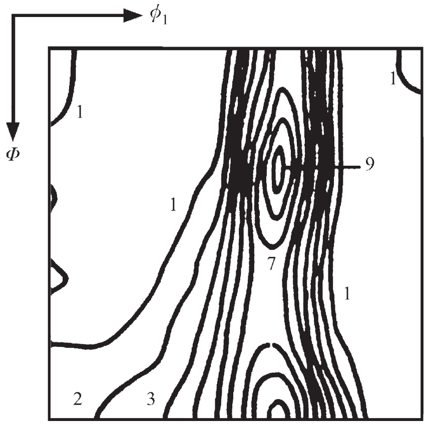


Figure 3
The $\phi_2=45^\circ$ section of an orientation distribution function for recrystallized decarburized steel cold rolled 70% and annealed (after Nes and Hutchinson 1989).

recrystallization and during grain growth, but the concentration of the main components may be changed. In copper, for example, low-temperature anneal leads to a relative strengthening of the [100]–[112] part of the deformation texture, whereas it is the [112]–[111] part that are the strongest after high-temperature recrystallization.

A complete change in texture during recrystallization is observed in titanium and zirconium wires. Here the $[10\bar{1}0]$ deformation texture is replaced by the $[11\bar{2}0]$ recrystallization texture.

A complete change in texture during grain growth is sometimes observed in tungsten wires where grains of the [320], [321], [531], or [421] orientations may grow abnormally, thus replacing the [110] recrystallization texture. However, these results vary according to tests by different observers and seem to depend critically upon alloy composition and presence of precipitates.

2. *Theories and Modeling Approaches to Simulate Annealing Textures*

2.1 Recrystallization

Two principally different mechanisms are generally used as the basis for recrystallization simulations: oriented nucleation and oriented growth. The oriented nucleation mechanism was suggested in the 1930s by W.G. Burgers and co-workers (e.g., Burgers and Louwerse 1931). According to this mechanism, preferred nucleation of grains with particular orientations occurs. Experimentally the importance of the oriented nucleation mechanism is checked by comparing the texture of an assembly of nuclei early in the recrystallization process to the texture in the fully recrystallized state. Often very good agreement has been observed. A range of theories has been put forward to explain and predict the orientation of these nuclei. The general belief is that the majority of the nuclei each have an orientation identical to that of the local area in the deformed microstructure where it forms. In metals with low to medium SFE where annealing twinning is frequent, many nuclei with orientations in twin relation (of order 1, 2, 3…) to the deformed matrix orientations are typically formed.

The orientated growth mechanism was suggested by P.A. Beck and others in the 1950s (e.g., Beck 1954). According to this mechanism, nuclei with certain preferred orientation relationships to the surrounding deformed matrix grow faster than other nuclei and thus contribute more strongly to the recrystallization texture. The most frequently observed preferential nucleus/matrix orientation relationships are in f.c.c. materials of 30–40° rotation about a common $\langle 111 \rangle$ direction, in b.c.c. materials 25–30° $\langle 110 \rangle$, and in c.p.h. materials 30° $\langle 0001 \rangle$. These preferential growth relationships can be very pronounced for tilt boundaries. The experimental verification for the oriented growth mechanism is provided mostly by controlled single-crystal or bicrystal experiments. For polycrystalline materials, the approach has been to transform the deformation texture by a preferential growth relation—e.g., 40° $\langle 111 \rangle$ including 1 to 8 of the eight possible $\langle 111 \rangle$ axes—and then compare this transformation texture to the recrystallization texture. Often very good agreement has been observed.

The relative importance of the oriented nucleation and oriented growth mechanisms has been a matter of controversy, but it is broadly accepted that both mechanisms can occur and that their relative importance depends on the specific material as well as deformation and recrystallization parameters. Besides this, a new trend in recrystallization modeling is to include both microstructural and orientational information. Challenges that are pursued are how to incorporate realistic deformation microstructures, including orientations, in the modeling and how to predict the correct nucleation.

2.2 Grain Growth

In spite of the experimental and theoretical evidence that texture and local orientation effects are important, such effects were only incorporated in the 1980s in grain-growth modeling. Both numerical and analytical models have been developed which can handle misorientation-dependent grain boundary energies and mobilities. Also drag effects of precipitates are considered and thus the combined effect of orientations and particles can be modeled. The models have been used for both normal and abnormal grain growth and it has been shown that orientation factors significantly affect the grain growth kinetics as well as the evolving microstructure. For more detailed information on texture and grain growth models see Srolovitz *et al.* (1984), Abbruzzese and Lücke (1986), and Novikov (1979).

Bibliography

Abbruzzese G, Lücke K 1986 A theory of texture controlled grain growth: I. Derivation and general discussion of the model. *Acta Metall.* **34**, 905–14

Barrett C, Massalski T B 1980 *Structure of Metals*, 3rd edn. Pergamon, Oxford, UK Chap. 21, pp. 568–83

Beck P A 1954 Annealing of cold worked metals. *Adv. Phys.* **3**, 245–324

Burgers W G, Louwerse P C 1931 Über den Zusammenhang zwischen Deformationsvorgang und Recrystallisationstextur bei Aluminium. *Zeit. Phys.* **67**, 605–17

Nes E, Hutchinson W B 1989 Texture and grain size control during processing of metals. In: Bilde-Sørensen J B, Hansen N, Juul Jensen D, Leffers T, Lilholt H, Pedersen O B (eds.) *Proc. 10th Risø Int. Symp. Materials Architecture*. Risø National Laboratory, Roskilde, Denmark, pp. 233–49

Novikov V Yu 1979 On computer simulation of texture development in grain growth. *Acta Metall.* **27**, 1461–6

Schmidt U, Lücke K 1979 Recrystallization textures of silver, copper and α-brasses with different zinc contents as a function of the rolling temperature. *Textures Cryst. Solids* **3**, 85–112

Srolovitz D J, Anderson M P, Grest G S, Sahni P S 1984 Computer simulation of grain growth: III. Influence of particle dispersion. *Acta Metall.* **32**, 1429–38

Wassermann G, Grewen J 1962 *Texturn Metallischer Werkstoffe*. Springer, Berlin

Wenk H-R (ed.) 1985 *Preferred Orientation in Deformed Metals and Rocks: An Introduction to Modern Texture Analysis*. Academic Press, Orlando, FL

D. J. Jensen
Risø National Laboratory, Roskilde, Denmark

Austenite Decomposition: Overall Kinetics during Isothermal, and Continuous Cooling Transformation

A detailed knowledge of the austenite decomposition kinetics in iron and steel can be of critical significance for the design or optimization of industrial processes to produce consistently high-quality steel. For example, the design of cooling regimes for hot strip, plate, and rod products must take into consideration the austenite decomposition since the resulting microstructure determines the mechanical properties of the steel. Consequently, process models are increasingly gaining attention in the steel industry with the goal being to link quantitatively the operational parameters of a mill with the properties of the steel product via accurate modeling of the microstructural evolution.

The austenite decomposition usually shows a sequential transformation pattern which may include the formation of the proeutectoid phase (either ferrite or cementite) followed by other transformations including pearlite, bainite, and martensite. Starting with perhaps the simplest single transformation case, the isothermal austenite-to-pearlite transformation in eutectoid steels, the model approaches will then be summarized for more complex multiphase situations resulting from continuous cooling transformations; the martensite transformation is not within the scope of this article.

1. Isothermal Austenite-to-pearlite Transformation

The effect of temperature on the austenite-to-pearlite transformation kinetics is shown in Fig. 1 for a eutectoid iron–carbon alloy containing 0.76 wt.%C. As the transformation temperature is decreased from 675 to 600 °C, the pearlite reaction is accelerated with both nucleation and growth rates increasing. Although many studies of the austenite-to-pearlite transformation in eutectoid carbon steels have been performed, significant variations in the kinetics due to differences in composition and/or austenite grain size preclude any absolute comparisons. For example, decreasing the austenite grain size results in a faster transformation reaction, the overall isothermal transformation curve in all cases being sigmoidal in shape. The Johnson–Mehl–Avrami–Kolmogorov (JMAK) analysis has been applied successfully to describe these transformation reactions on a semiempirical level. A generalized form of the JMAK analysis was proposed by Avrami (1940) and is now

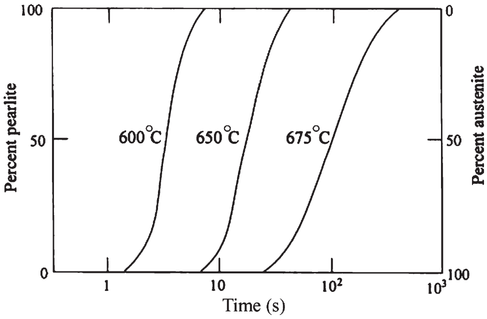


Figure 1
Kinetics of the isothermal austenite-to-pearlite transformation for a eutectoid Fe–C alloy containing 0.76 wt.%C (after Callister 1985).

commonly known as the Avrami equation; i.e.,

$$f = 1 - \exp(-bt^n) \tag{1}$$

where f is the fraction transformed, t is the time at temperature, and b and n are rate constants to be determined from experimental data. The value of n can usually be taken as being independent of temperature. Umemoto *et al.* (1980) reported an n value of 4 obtained by including the incubation time in the isothermal transformation analysis.

Considering only the transformation time for 1% to 99% completion, Hawbolt *et al.* (1983) determined a value of approximately 2 for n. The parameter b increases with decreasing temperature, T, and $T \geqslant 600\,^\circ\mathrm{C}$ because of increased transformation rates. Umemoto *et al.* (1980) incorporated into the Avrami equation the effect of austenite grain size on the transformation kinetics by expressing b as a function of the initial austenite grain size, d_γ, such that

$$b = \frac{b_0(T)}{d_\gamma^m} \tag{2}$$

where b_0 is a function of temperature and m is the grain size exponent. The value of m indicates the operational nucleation mode with $m = 1$ for nucleation at grain faces, $m = 2$ for nucleation at edges, and $m = 3$ for nucleation at corners. For the pearlite reaction, $m = 1.76$ was determined, indicating predominant nucleation at grain edges. An n value of 2, as obtained for the 1% to 99% transformation period, is then consistent with nucleation site saturation.

The austenite-to-pearlite reaction, which occurs in noneutectoid steels after the formation of the proeutectoid phase, can be described in a very similar way with the help of the Avrami equation by renormalizing the pearlite portion of the transformation to unity. Campbell *et al.* (1991) analyzed the pearlite reactions for different chemistries of hypoeutectoid steels with carbon concentrations in the range 0.2–0.7 wt.%. Analogous to eutectoid steels, n varies little with temperature, whereas b increases with undercooling. Analyzing the 1 to 99% transformation period only, the n value decreases from approximately 2 to 1 as the carbon concentration, C, decreases from 0.8 to 0.2 wt.%, such that

$$n = 3.01C^2 - 1.06C + 0.5Mn + 0.792 \tag{3}$$

where the concentrations for carbon and manganese are in wt.%. The decrease in n with decreasing carbon content can probably be attributed to the change in location of pearlite nucleation from austenite grain boundaries to ferrite–austenite (α–γ) interfaces.

The Avrami equation is a powerful tool to describe quantitatively the transformation kinetics, but gives little insight into the actual transformation mechanisms. The rate constants b or b_0 are frequently expressed as an empirical function of undercooling (Campbell *et al.* 1991). Other researchers (Suehiro *et al.* 1987, Umemoto *et al.* 1992) quantified these rate constants by employing explicitly nucleation and growth rates. The nucleation rate can be described with classical nucleation theory. The pearlite growth rate was assumed to be that proposed by Zener (1946) and Hillert (1957) for volume diffusion controlled by carbon in austenite.

2. *Formation of Proeutectoid Phases*

In steels which are not of eutectoid composition, the pearlite transformation occurs after the formation of the proeutectoid phase—ferrite or cementite. The austenite-to-ferrite transformation has received much more attention since the majority of steels are of hypoeutectoid composition. As a result, modeling of the overall transformation kinetics has concentrated on the formation of proeutectoid ferrite.

The transformation curve of ferrite is quite different from that of pearlite. There is a clear incubation time, t_0, for the initiation of ferrite formation and the transformation rate exhibits its maximum at the beginning of the reaction. However, by normalizing the ferrite fraction to one, the Avrami equation can be applied when the time, t, in Eqn. (1) is replaced by $t - t_0$. Similar to the pearlite reaction, n may be taken as independent of temperature and b can be expressed as a function of temperature and austenite grain size in the form of Eqn. (2). As a general rule, $n = 1$ and $m = 1$ may be taken assuming nucleation site saturation at austenite grain boundaries. For low- and medium-carbon steels, the experimentally determined n values range from 0.88 to 1.33 (Campbell *et al.* 1991). Umemoto *et al.* (1984) determined m to be approximately 1.2–1.3, which is consistent with predominant nucleation at grain faces at higher undercoolings (Lange and Aaronson 1979, Enomoto *et al.* 1986).

Further, (Umemoto *et al.* 1984) found that the Avrami equation predicted faster transformation rates in the later transformation stages. This discrepancy can be attributed to the overlapping of the diffusion fields from adjacent ferrite grains. By multiplying the transformation rate from the Avrami equation by $(1-X)^p$ where $p = 0.5$, the following equation for the normalized ferrite fraction transformed was obtained

$$f = 1 - \left[1 + \frac{0.5b_0(T)}{d_\gamma^m}\{t - t_0(T)\}^n\right]^{-2} \tag{4}$$

to express better the overall transformation kinetics. The incubation time is determined by the nucleation rate of ferrite which can be described with classical nucleation theory (see *Precipitation from Austenite*). Vandermeer (1992) proposed a more sophisticated consideration of the nucleation by incorporating a microstructural path function. This approach requires the numerical solution of coupled equations for the fraction transformed and the α–γ interfacial area per unit volume. In the limits of short and long transformation times, Avrami-type equations are obtained with $n = 5/2$ and $n = 1/2$, respectively.

In the models assuming nucleation site saturation, the rate constant b is determined by subsequent ferrite growth. Empirical expressions have been proposed representing b as a function of temperature (Campbell *et al.* 1991). In the overall transformation models of Suehiro *et al.* (1987) and Umemoto *et al.* (1992), rate expressions are formulated based on fundamental growth theories. In Fe–C alloys, growth of the ferrite allotriomorphs is controlled by volume diffusion of carbon in austenite and their half thickness, S, is given by

$$S = \alpha\sqrt{t} \tag{5}$$

where α is the parabolic growth constant which can be determined from a solution of the one-dimensional diffusion problem by assuming a weighted average diffusivity of carbon in austenite (Atkinson *et al.* 1973). Umemoto *et al.* (1992) employed a simple approximation of α which is based on the linearized concentration profile introduced by Zener (1949). Suehiro *et al.* (1987) took a somewhat different approach, adopting the equation of Zener (1946) and Hillert (1957) for volume diffusion-controlled plate lengthening, applicable to the growth kinetics of Widmanstätten ferrite. Further, in alloys containing substitutional alloying elements, solute drag-like effects (SDLE) may reduce growth rates, as shown by Bradley and Aaronson (1981) for ternary alloys such as Fe–C–Mn and Fe–C–Cr. In particular, the SDLE of manganese are of practical importance since most steels contain manganese; plain carbon steels are essentially Fe–C–Mn alloys.

In addition to the Avrami-type overall transformation models, fundamental diffusion models were developed for the austenite-to-ferrite transformation. Ferrite growth is parabolic as long as overlapping of diffusion fields from neighboring ferrite grains does not affect the transformation rate. To account for the overall transformation kinetics, numerical solutions of the diffusion problem were proposed (Vandermeer 1990, Kamat *et al.* 1992, Silalahi *et al.* 1995). The ferrite diffusional growth process is described by assuming nucleation site saturation at austenite grain boundaries, α–γ interfacial equilibrium, a spherical austenite grain having a diameter equal to the initial austenite grain size, and no carbon flux at the center of the spherical grain. Further, the model of Vandermeer (1990) incorporates a log-normal austenite grain size distribution for polycrystalline material. These diffusion models give good agreement with experimental observations of isothermal transformation kinetics in Fe–C alloys and low-manganese, plain carbon steels.

3. *Formation of Bainite*

From a kinetic point of view, bainite forms when the steel is cooled well below the eutectoid temperature to the upper limiting temperature, B_s, where sympathetic nucleation (SN) of ferrite at the α–γ interface becomes significant. For the bainite transformation kinetics, a number of theories has been proposed ranging from displacive, diffusionless transformation to ledgewise diffusional growth (see *Bainite*). Aside from the controversy regarding the growth mechanisms, the Avrami model has been applied to describe empirically the bainite kinetics. However, different rate constants were reported and the procedure appears to be limited since the variations in the rate constants cannot be predicted. For example, Umemoto *et al.* (1982) determined $n = 4.8$ and $m = 0.65$ for a high-carbon steel containing 0.99 wt.%C and 1.39 wt.%Cr. These values for n and m are consistent with the occurrence of SN; i.e., nucleation and growth take place simultaneously at α–γ boundaries with nucleation occurring also in the grain interior ($m < 1$). In the overall transformation models of Suehiro *et al.* (1987) and Umemoto *et al.* (1992), the rate constant b is described by adopting a diffusional transformation model for bainite. In contrast, Rees and Bhadeshia (1992) developed an overall bainite transformation model for steels with more than 1.5 wt.%Si, where carbide precipitation is suppressed, by assuming a diffusionless displacive mechanism in which bainite grows through successive nucleation of subunits. Minote *et al.* (1996) suggested for steels containing approximately 0.2 wt.%C, 1.5 wt.%Mn, and 1.5 wt.%Si that the approach assuming a diffusion mechanism gives more accurate predictions at temperatures above 350 °C, whereas at temperatures below 350 °C the model based on a displacive mechanism appears to be more suitable.

4. *Continuous Cooling Transformation*

Modeling the transformation kinetics can be extended from the isothermal to nonisothermal conditions by utilizing the additivity rule for individual transformation products as proposed by Cahn (1956). The additivity rule states that if $\tau(f,T)$ is the time required to reach the transformed fraction, f, at a constant temperature, T, starting from $f = 0$, f is attained in an anisothermal process after time, t, such that

$$\int_0^t \frac{\mathrm{d}t'}{\tau[f, T(t')]} = 1 \tag{6}$$

The transformation is called isokinetic when the transformation rate can be written in a factorized form:

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \xi(f)\zeta(T) \tag{7}$$

An isokinetic transformation fulfils the additivity rule (Lusk and Jou 1997). Further, Cahn (1956) stated that the rule of additivity is valid for any rate-independent transformation, i.e., where the transformation rate only depends on the fraction transformed and temperature. This is not true in general, as shown by Lusk and Jou (1997). A nonfactorized, rate-independent reaction may not exactly fulfil Eqn. (6), but the deviations appear to be comparatively minor. Furthermore, integration along the thermal path is always possible for any rate-independent transformation. The individual transformation products (e.g., ferrite, pearlite, bainite) considered here can usually be described with an Avrami model where n is temperature-independent. Then the transformation rate fulfils Eqn. (7) and the additivity rule can be applied.

The fraction transformed at temperature T can be written as

$$f = 1 - \exp\left\{ -\frac{1}{d_\gamma^m} \left(\int_T^{T_\mathrm{e}} \frac{b_0(T)^{1/n}}{\phi(T)} \mathrm{d}T \right)^n \right\} \tag{8}$$

where T_e is the temperature below which there is a driving force for the transformation and $\phi = \mathrm{d}T/\mathrm{d}t$ is the instantaneous cooling rate. Drawing on the additivity concept, Suehiro *et al.* (1987) and Umemoto *et al.* (1992) extended their overall transformation models to continuous cooling. Furthermore, the model of Umemoto *et al.* (1992) incorporates criteria for the transition between the sequential formation of different transformation products, as well as the transformation from work-hardened austenite, which is relevant for niobium microalloyed steels.

Militzer *et al.* (2000) developed a process model for the austenite-to-ferrite transformation during continuous cooling of low-carbon steels including microalloyed steels. The model consists of three parts: (i) prediction of the transformation start temperature, T_S; (ii) ferrite growth model; and (iii) ferrite grain size prediction. Considering carbon diffusion controlled early growth of ferrite nucleated at grain corners the transformation start may be calculated from

$$c^* - c^o = \frac{2(c_\gamma - c^o)}{\phi^{1/2} d_\gamma} \sqrt{\int_{T_\mathrm{S}}^{T_\mathrm{N}} D_\mathrm{C} \frac{c_\gamma - c^o}{c_\gamma - c_\alpha} \mathrm{d}T} \tag{9}$$

where c^o is the bulk carbon concentration, c_α and c_γ are the equilibrium carbon concentrations in the ferrite and austenite, respectively, D_C is the carbon diffusion coefficient in austenite, and ϕ is the average cooling rate between the nucleation temperature, T_N, and T_S. The parameter c^* denotes a critical carbon concentration above which ferrite nucleation is inhibited, with c^* being approximately 1.3 c^o for plain carbon steels.

Subsequent ferrite growth can be described either with Eqn. (8), where T_e is replaced by T_S, and $n = 0.9$ is adopted or with a rigorous diffusion model. Militzer *et al.* (1996) modified the carbon diffusion model by incorporating the SDLE of manganese with a phenomenological segregation factor

$$s = \exp(G(T)/RT) \tag{10}$$

where G is an effective segregation energy which decreases with undercooling. Increasing the cooling rate decreases the transformation temperature, as shown for a 0.038 wt.%C–0.3 wt.%Mn plain carbon steel in Fig. 2; the model predictions are in good agreement with the experimental data.

For interstitial-free (IF) steels, the above diffusion model is not applicable since the austenite-to-ferrite transformation is interface-controlled. Manganese, and possibly niobium, exert a similar SDLE at the moving α–γ interface, as discussed for low-carbon steels. The growth rate of ferrite may then be calculated from the intrinsic velocity of the α–γ interface in iron which is reduced by an appropriate segregation factor, s (Militzer 1999).

Early transformation stages determine the ferrite grain size, d_α, which can be predicted as a function of the transformation start temperature, such that

$$d_\alpha = \left\{ B(d_\gamma)\exp(-E/T_\mathrm{S})F \right\}^{1/3} \tag{11}$$

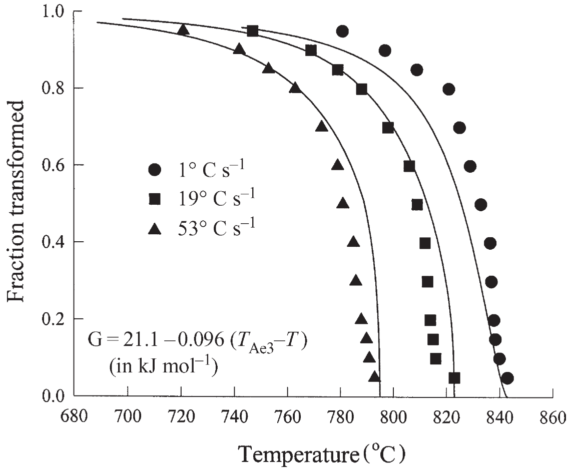


Figure 2
Continuous cooling transformation kinetics in a 0.038–0.30 wt.%Mn plain carbon steel; symbols indicate experimental results, solid lines give predictions from carbon diffusion-controlled ferrite growth modified by interfacial manganese segregation (after Militzer *et al.* 1996).

where T_S is in K and F is the final ferrite fraction. The parameters B, which depends on the initial austenite microstructure, and E, are determined experimentally.

The prediction of the final ferrite fraction can be a challenging task for the wide range of steel chemistries and processing conditions which often results in multiphase structures. For carbon–manganese steels, the transition from ferrite to pearlite can be predicted with a temperature-dependent critical velocity of the moving α–γ interface, below which the nucleation of cementite occurs (Umemoto *et al.* 1992).

A more complex situation arises when two or more transformation products form simultaneously. Jones and Bhadeshia (1997) proposed an empirical description based on an extension of the Avrami model where, in general, a system of coupled differential equations for the rates of each individual transformation product has to be solved numerically. Detailed investigations of the complex phase transformation kinetics in multiphase steels are an area of current research with the goal being to design processing routes for steels with enhanced properties (Minote *et al.* 1996).

Bibliography

Atkinson C, Aaron H B, Kinsman K R, Aaronson H I 1973 On the growth kinetics of grain boundary ferrite allotriomorphs. *Metall. Trans.* **4**, 783–92

Avrami M 1940 Kinetics of phase change II. *J. Chem. Phys.* **8**, 212–24

Bradley J R, Aaronson H I 1981 Growth kinetics of grain boundary ferrite allotriomorphs in Fe–C–X alloys. *Metall. Trans. A* **12**, 1729–41

Cahn J W 1956 Transformation kinetics during continuous cooling. *Acta Metall.* **4**, 572–5

Callister W D Jr 1985 *Materials Science and Engineering: An Introduction*. Wiley, New York

Campbell P C, Hawbolt E B, Brimacombe J K 1991 Microstructural engineering applied to the controlled cooling of steel wire rod: Part II. Microstructural evolution and mechanical properties correlations. *Metall. Trans. A* **22**, 2779–90

Enomoto M, Lange W F III, Aaronson H I 1986 The kinetics of ferrite nucleation at austenite grain edges in Fe–C and Fe–C–X alloys. *Metall. Trans. A* **17**, 1399–407

Hawbolt E B, Chau B, Brimacombe J K 1983 Kinetics of austenite–pearlite transformation in eutectoid carbon steel. *Metall. Trans. A* **14**, 1803–15

Hillert M 1957 The role of interfacial energy during solid state phase transformations. *Jernkontorets Ann.* **141**, 757–89

Jones S J, Bhadeshia H K D H 1997 Kinetics of the simultaneous decomposition of austenite into several transformation products. *Acta Mater.* **45**, 2911–20

Kamat R G, Hawbolt E B, Brown L C, Brimacombe J K 1992 The principle of additivity and the proeutectoid ferrite transformation. *Metall. Trans. A* **23**, 2469–80

Lange W F III, Aaronson H I 1979 Effect of undercooling upon relative nucleation kinetics of proeutectoid ferrite at austenite grain faces and grain edges. *Metall. Trans. A* **10**, 1951–2

Lusk M, Jou H J 1997 On the rule of additivity in phase transformation kinetics. *Metall. Mater. Trans. A* **28**, 287–91

Militzer M 1999 Austenite decomposition kinetics in advanced low carbon steels. In: Koiwa M, Otsuka K, Miyazaki T (eds.) *Solid–Solid Phase Transformations, Proceedings Volume 12*. The Japan Institute of Metals, Sendai, pp. 1521–4

Militzer M, Hawbolt E B, Meadowcroft T R 2000 Microstructural model for hot strip rolling of high-strength low-alloy steels. *Metall. Mater. Trans. A* **31**, 1247–59

Militzer M, Pandi R, Hawbolt E B 1996 Ferrite nucleation and growth during continuous cooling. *Metall. Mater. Trans. A* **27**, 1547–56

Minote T, Torizuka S, Ogawa A, Niikura M 1996 Modeling of transformation behavior and compositional partitioning in TRIP steel. *ISIJ Int.* **36**, 201–7

Rees G I, Bhadeshia H K D H 1992 Bainite transformation kinetics: Part 1. Modified model. *Mater. Sci. Technol.* **8**, 985–93

Silalahi V M M, Onink M, van der Zwaag S 1995 Decomposition of hypo-eutectoid Fe–C austenites; a numerical diffusion model. *Steel Res.* **66**, 482–9

Suehiro M, Sato K, Tsukano Y, Yada H, Senuma T, Matsumura Y 1987 Computer modeling of microstructural change and strength of low-carbon steel in hot strip rolling. *Trans. ISIJ* **27**, 439–45

Umemoto M, Hiramatsu A, Moriya A, Watanabe T, Nanba S, Nakajima N, Anan G, Higo Y 1992 Computer modelling of phase transformation from work-hardened austenite. *ISIJ Int.* **32**, 306–15

Umemoto M, Horiuchi K, Tamura I 1982 Transformation kinetics of bainite during isothermal holding and continuous cooling. *Trans. ISIJ* **22**, 854–61

Umemoto M, Komatsubara N, Tamura I 1980 Prediction of hardenability effects from isothermal transformation kinetics. *J. Heat Treat* **1**, 57–64

Umemoto M, Nishioka N, Tamura I 1984 Kinetics of proeutectoid ferrite reaction during isothermal holding and continuous cooling in plain carbon steels. In: Bell T (ed.) *Heat Treatment Shanghai 1983*, The Metals Society, London, pp. 5.35–43

Vandermeer R A 1990 Modeling diffusional growth during austenite decomposition to ferrite in polycrystalline Fe–C alloys. *Acta Metall. Mater.* **38**, 2461–70

Vandermeer R A 1992 Microstructural modeling of austenite decomposition to ferrite in an HSLA steel. In: DeArdo A J (ed.) *Proc. Int. Conf. Processing, Microstructure and Properties of Microalloyed and Other Modern High Strength Low Alloy Steels*, ISS, Warrendale, PA, pp. 33–9

Zener C 1946 Kinetics of the decomposition of austenite. *Trans. AIME* **167**, 550–83

Zener C 1949 Theory of growth of spherical precipitates from solid solution. *J. Appl. Phys.* **20**, 950–3

M. Militzer
University of British Columbia, Vancouver, British Columbia, Canada

This Page Intentionally Left Blank

B

Bainite

Bainite is a microstructural product of eutectoid decomposition; it forms when a high-temperature phase decomposes into two different phases on cooling. It is distinguished from pearlite, another eutectoid decomposition product, by its morphology. Bainite is a nonlamellar mixture of the two product phases that evolves when these phases grow at different rates. Pearlite is a lamellar mixture that forms when the product phases grow cooperatively. Bainitic microstructures have significant commercial value because they can be produced economically with simple processing, and they provide some alloy steels and cast irons with an attractive combination of mechanical properties.

Although bainite is found in many nonferrous alloys, research on the subject since the early 1900s has focused primarily on alloy steels. A brief description of some of the intriguing phenomena associated with bainite in ferrous alloys is provided with an emphasis on the physical metallurgy of bainite. A more complete introduction to the subject is available in several publications (Bainite 1990, 1991, Pacific Rim 1994, Hillert 1995).

1. Introduction

The first systematic investigations of the intermediate product, now called bainite in honor of Edgar C. Bain, began with observations of steel microstructures produced by isothermal decomposition of austenite (Hultgren 1920, Robertson 1929, Davenport and Bain 1930). These workers used optical microscopy to show bainite (i) has an acicular structure, (ii) contains carbides, and (iii) appears mainly at temperatures below the range for pearlite. Early work showed it is found in many alloy steels and its appearance varies considerably with temperature and composition (Griffiths *et al.* 1939, Hultgren 1947).

The first data on the kinetics of bainite formation came from Germany. Magnetization and dilatometry measurements on alloy steels obtained during austenite decomposition show bainite forms rapidly at first but then stops well before the parent austenite is completely consumed (Wever and Lange 1932, Wever and Jellinghaus 1932). This behavior, which is limited to the higher temperatures at which bainite forms, is called the incomplete reaction phenomenon or transformation stasis. The amount of bainite present when transformation ceases increases from zero at the uppermost temperature of bainite formation (designated the bainite start or B_s temperature) to nearly 100% as the transformation temperature decreases.

Based on the etching characteristics of bainite, early workers speculated the ferritic component of bainite incorporates an increasing amount of carbon from austenite as the transformation temperature decreases. However, it was also clear that bainite formation kinetics exhibit characteristics of a nucleation and growth process and differ drastically from martensite formation kinetics.

2. Displacive Mechanism

Proposed mechanisms for the bainite reaction have evolved along two lines of reasoning: the "displacive" hypothesis that emphasizes similarities between bainite and martensite, and the "diffusional" hypothesis that emphasizes commonalties between bainite and proeutectoid ferrite. The displacive model, first articulated by Robertson and by Davenport and Bain and later developed by others, envisioned bainite formation as a two-step process: an essentially martensitic change of crystal structure that produces carbon supersaturated ferrite followed almost immediately by carbide precipitation (Robertson 1929, Davenport and Bain 1930). Early versions of this theory supposed the ferrite forms in carbon-poor regions of austenite, with the ferritic component inheriting the carbon concentration of the carbon-poor austenite. However, carbon-poor regions in austenite on the size scale of bainite crystals have not been observed experimentally and they do not appear plausible on statistical grounds, so this aspect of the theory was eventually discarded.

Zener provided a thermodynamic basis for the displacive hypothesis in a seminal paper on ferrous transformations published in 1946. His model assumed the ferritic component of bainite (i.e., bainite before carbides appear) inherits the full carbon concentration of the parent austenite. This approach identified the B_s temperature as the critical temperature below which it is possible to convert austenite to ferrite without a composition change. The model distinguished bainitic ferrite from martensite by supposing bainite forms as a strain-free product while martensite forms with accompanying strain energy.

Zener suggested transformation stasis could arise from a loss of driving force for ferrite growth as carbon leaks out of ferrite and into the surrounding austenite. As was then noted in the published discussion accompanying Zener's paper (Troiano 1946), this explanation for stasis is not satisfactory because it requires a change in the ferrite composition during growth, thereby violating Zener's initial assumption that ferrite and austenite have identical compositions. It was also noted that the calculated B_s temperatures did not agree with existing data. Predicted B_s

temperatures in Fe–C–Mo and Fe–C–Cr alloys have been shown to lie as much as 200 °C above actual bainite start temperatures (Enomoto and Tsubakino 1991). More sophisticated displacive models for bainite growth include strain energy contributions to bainite and incorporate partial carbon supersaturation in ferrite during growth (Hsu 1990, Bhadeshia and Christian 1990, Olson *et al.* 1990). However, these models remain to be tested against available data from a series of alloys.

Attempts have also been made to determine the carbon concentration in bainitic ferrite before carbides appear, but most have relied upon indirect methods or yielded a wide range of concentrations (Aaronson *et al.* 1990). This important question remains to be resolved.

In 1952 Ko and Cottrell made observations of polished surfaces subsequently transformed to bainite to show bainite produces a surface relief similar to that of martensite. They also employed hot-stage microscopy to follow bainite formation *in situ* and found bainite grows much more slowly than martensite. They reasoned bainitic ferrite grows by a martensitic mechanism but at a rate paced by the removal of carbon from ferrite. Carbon removal was suggested to occur by diffusion into austenite, by precipitation of carbides in ferrite, or both. In contrast to Zener's approach, Ko and Cottrell argued the carbon had to be removed from ferrite during growth (rather than after growth) in order to reduce the dilatational strain when ferrite forms and to provide sufficient driving force to overcome the shear strain associated with a martensitic transformation mechanism. The finding that bainite grows slowly was verified by bainite lengthening and thickening kinetics measurements made later with other techniques—the measured rates were often slower than rates calculated on the basis of carbon diffusion in austenite (Aaronson and Hall 1994, Hillert 1995).

The suggestion that the bainite structure change is essentially martensitic has not been supported by experiment. Unlike martensite, whose interfacial structure is glissile (Sandvik and Wayman 1983), carbide-free bainite has a sessile interfacial structure that is incapable of moving in martensitic fashion (Rigsbee and Aaronson 1979, Li *et al.* 1988). In addition, single-crystal plates of ferrite have been shown to produce tent-shaped surface reliefs—a type of relief wholly incompatible with a martensitic transformation mechanism (Hall and Aaronson 1994).

3. Diffusional Mechanism

Following a systematic investigation of microstructures in a series of alloy steels, Hultgren proposed a diffusional hypothesis for bainite formation. He suggested bainite evolved with the nucleation of probainitic (acicular) ferrite, then nucleation of cementite particles on the ferrite/austenite boundaries, followed by the cementite being surrounding by growing ferrite, and eventually nucleation of new cementite particles. In contrast to bainite, pearlite was suggested to develop by nucleation of a cementite plate, followed by an adjacent ferrite plate, and then coordinated lengthening of the plates.

Electron microscopy observations of bainite and pearlite have led to a mechanistic basis for this approach. The resulting, more detailed, model articulates the circumstances under which bainite and pearlite form during a general eutectoid reaction (Lee *et al.* 1988). This model relates the internal morphology of a eutectoid product to two factors: (i) the relative growth rates of the eutectoid phases, and (ii) the rate at which the slower-growing phase renucleates on the boundary between the faster-growing phase and the parent (Fig. 1). An attractive feature of the model is its ability to explain a variety of bainite morphologies (Fig. 2) in terms of simple factors that, at least in principle, can be measured or estimated. For example, the model has been applied to interpret transitions in internal morphology and in aggregate shape of evolving eutectoid structures (the external morphology) as a function of transformation temperature and carbon concentration in a series of Fe–C–2 wt.% Mn alloys (Spanos *et al.* 1990a, b). By considering how the nucleation and growth rates of ferrite and cementite change with temperature and carbon concentration, it is possible to explain the relatively sharp transition between upper and lower bainite and the more gradual changes from upper and lower bainite to nodular bainite.

Hultgren's original model of bainite formation also proposed an explanation for transformation stasis

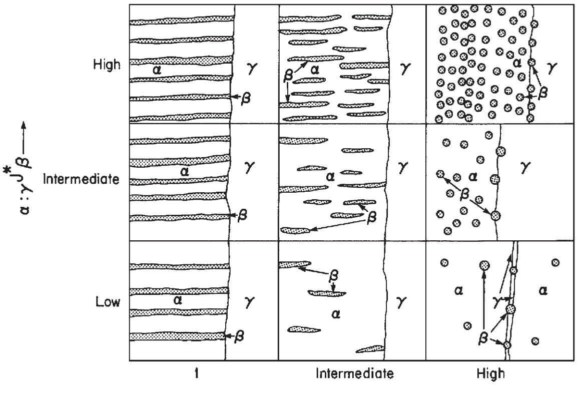


Figure 1
Variation in eutectoid microstructures with ratio of the product growth rates, G_α/G_β, and nucleation rate of the slower growing product, ${}_{\alpha-\gamma}J^*_\beta$, when the volume fraction of the slower growing phase is small (after Lee *et al.* 1988).

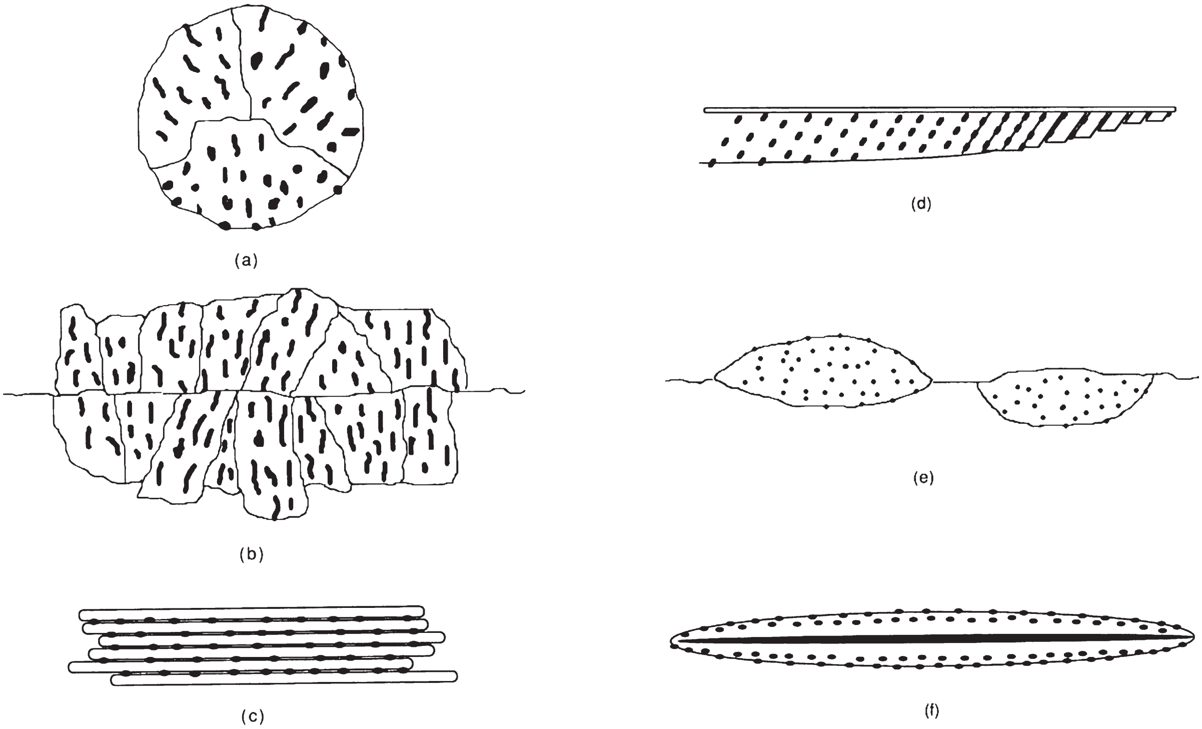


Figure 2
Schematic illustrations of various bainite morphologies. The white constituent represents the majority eutectoid phase (e.g., ferrite) and the dark constituent represents the minority phase (e.g., cementite). (a) Nodular bainite, (b) columnar bainite along a prior matrix grain boundary, (c) a sheaf of upper bainite laths, (d) lower bainite, (e) grain boundary allotriomorphic bainite, and (f) inverse bainite (after Spanos *et al.* 1990b).

(Hultgren 1947). Two types of ferrite were envisioned: orthoferrite, which differs from austenite in carbon concentration and in the concentrations of substitutional alloying elements, and paraferrite (paraequilibrium ferrite), which differs from austenite in carbon concentration but not in alloy content. It was suggested that the B_s temperature is the temperature below which paraferrite appears, and transformation stasis just below the B_s temperature is caused by a slow reaction rate associated with the upper portion of the "C" curve of paraferrite on the TTT diagram. This explanation for the B_s temperature and transformation stasis does not explain subsequent experimental results. Partitioning data and measured growth kinetics from a variety of Fe–C–X alloys indicate the transition to paraferrite occurs at temperatures well above the range of observed B_s temperatures (Aaronson *et al.* 1988). Also, the amount of ferrite found during transformation stasis in Fe–C–Cr and Fe–C–Mo alloys is well below the metastable fraction expected to form under paraequilibrium (Enomoto and Tsubakino 1991).

A promising framework for understanding transformation stasis was built on the effects of alloying elements on ferrite growth. Measured transformation kinetics in ternary alloys (e.g., Reynolds *et al.* 1990a, b) have shown transformation stasis is not a general property of bainite, but rather a characteristic of particular alloying elements and concentrations. In general, alloys that exhibit transformation stasis also tend to be those in which ferrite growth is significantly slower than predicted under paraequilibrium (Shiflet and Aaronson 1990). The ability of specific solute species to slow ferrite growth is widely attributed to a solute drag effect, or solute drag-like effect, depending on how one describes the origin of the drag force (Reynolds *et al.* 1990a, Purdy and Brechet 1995, Enomoto 1999). Efforts to ascertain the degree of solute adsorption at ferrite/austenite boundaries have yielded somewhat mixed results, although the presence of molybdenum accumulations at these boundaries in Fe–C–Mo alloys now appears certain (Aaronson *et al.* 1990, Aaronson *et al.* in press). There is, however, a large body of microstructural and kinetic evidence for the solute drag effect in steels, and investigations into interactions between solutes and growth interfaces are likely to be a productive area for research.

The central assumption needed to explain transformation stasis below B_s is that the solute drag effect must be strong enough to slow markedly or stop ferrite growth. This is feasible when sufficient concentrations of carbon and of an effective solute are present, the reaction temperature is relatively low, and ferrite boundaries have migrated far enough to pick up the dragging solute. Under these circumstances, renucleation of ferrite on immobilized ferrite/austenite boundaries (sympathetic nucleation) becomes feasible if sufficient undercooling is available. The temperature below which copious sympathetic nucleation begins corresponds to the kinetic B_s temperature. The degenerate ferrite morphology found below the bay temperature results from a change from relatively smooth, ledgewise migration of ferrite boundaries to irregular "start–stop" migration during which ferrite nucleates, grows until stopped by solute drag, and is then forced to renucleate again (Reynolds *et al.* 1990a).

4. Summary

Hypotheses for the mechanism of bainite formation in steels have been discussed in the context of available experimental evidence. Single-crystal ferrite plates often exhibit complex surface reliefs that are inconsistent with a displacive transformation mechanism. The interfacial structure found on ferrite plates formed at low temperatures is sessile and must migrate in a nonconservative fashion. Also, B_s temperatures calculated assuming a displacive growth mechanism for ferrite are in poor agreement with the B_s temperatures of many steels. These circumstances indicate the ferrite component of bainite does not form by a displacive transformation mechanism.

Features of the internal and external morphologies of bainite can be explained by considering the relative nucleation rates and ledgewise, diffusional growth rates of the product phases. Measured ferrite growth kinetics significantly slower than those calculated assuming paraequilibrium are found in alloys that exhibit transformation stasis. Slow ferrite growth in these alloys is ascribed to a solute drag effect. When potent, the solute drag effect is believed to be responsible for the appearance of copious sympathetic ferrite nucleation below B_s. If this initially takes place without significant interphase boundary carbide precipitation, transformation stasis appears.

See also: Pearlite

Bibliography

Aaronson H I, Enomoto M, Reynolds Jr. W T 1988 Influence of the chemistry and structure of austenite:ferrite boundaries upon growth kinetics and composition of grain boundary ferrite allotriomorphs in Fe–C–X alloys. In: Embury J D, Purdy G R (eds.) *Advances in Phase Transitions*. Pergamon, New York, pp. 20–36

Aaronson H I, Fletcher H A, Smith G D W, Garratt-Reed A J, Reynolds Jr. W T, Purdy G R 2000 Effects of Mo on ferrite formation in Fe–C–Mo alloys. *Steel Heat Treating in the New Millennium*. ASM International, Materials Park, OH

Aaronson H I, Hall M G 1994 A history of the controversy over the roles of shear and diffusion in plate formation above Md and a comparison of the atomic mechanisms of these processes. *Metall. Mater. Trans.* **25A**, 1797–819

Aaronson H I, Reynolds Jr. W T, Shiflet G J, Spanos G 1990 Bainite viewed three different ways. *Metall. Trans.* **21A**, 1343–80

Bainite 1990 Symposium on international conference on bainite: parts I and II. *Metall. Trans.* **21A**, 765–893, 1343–556

Bainite 1991 Special issue on the bainitic transformation. *Mater. Trans. JIM* **32**, 631–789

Bhadeshia H K D H, Christian J W 1990 On bainite formation. *Metall. Trans.* **21A**, 767–97

Davenport E S, Bain E C 1930 Transformation of austenite at constant subcritical temperatures. *Trans. AIME* **90**, 117–54

Enomoto M 1999 Influence of solute drag on the growth of proeutectoid ferrite in Fe–C–Mn alloy. *Acta Metall.* **47**, 3533–40

Enomoto M, Tsubakino H 1991 Morphology and thermodynamics of bainitic transformation in ferrous and non-ferrous alloys. *Mater. Trans. JIM* **32**, 642–57

Griffiths W T, Pfeil L B, Allen N P 1939 The intermediate transformation in alloy steels. *Second Report Alloy Steels Research Committee*, Special Report No. 24. Iron and Steel Institute, London, pp. 343–67

Hall M G, Aaronson H I 1994 Formation of invariant plane-strain and tent-shaped surface reliefs by the diffusional ledge mechanism. *Metall. Mater. Trans.* **25A**, 1923–31

Hillert M 1995 The nature of bainite. *ISIJ Int.* **35**, 1134–40

Hsu T Y 1990 On bainite formation. *Metall. Trans.* **21A**, 811–16

Hultgren 1920 *A Metallographic Study on Tungsten Steels*. Wiley, New York

Hultgren A 1947 Isothermal transformation of austenite. *Trans. ASM* **39**, 915–89

Ko T, Cottrell S A 1952 The formation of bainite. *JISI* **172**, 307–13

Lee H J, Spanos G, Shiflet G J, Aaronson H I 1988 Mechanisms of the bainite (non-lamellar eutectoid) reaction and a fundamental distinction between bainite and pearlite (lamellar eutectoid) reactions. *Acta Metall.* **36**, 1129–40

Li C, Perovic V, Purdy G R 1988 Observations of the bainite/austenite interface in steels. In: Lorimer G W (ed.) *Phase Transformations '87*. Institute of Metals, London, pp. 326–9

Olson G B, Bhadeshia H K D H, Cohen M 1990 Coupled diffusional/displacive transformations: II. Solute trapping. *Metall. Trans.* **21A**, 805–9

Pacific Rim 1994 Proceedings of the Pacific Rim conference on the roles of shear and diffusion in the formation of plate-shaped transformation products. *Metall. Trans.* **25A**, 1758–2016, 2553–673

Purdy G R, Brechet Y J M 1995 A solute drag treatment of alloying elements on the rate of the proeutectoid ferrite transformation in steels. *Acta Metall.* **43**, 3763–74

Reynolds Jr. W T, Li F Z, Shui C K, Aaronson H I 1990a The incomplete transformation phenomenon in Fe–C–Mo alloys. *Metall. Trans.* **21A**, 1433–63

Reynolds Jr. W T, Liu S K, Li F Z, Hartfield S, Aaronson H I 1990b An investigation of the generality of incomplete transformation to bainite in Fe–C–X alloys. *Metall. Trans.* **21A**, 1479–91

Rigsbee J M, Aaronson H I 1979 The interfacial structure of the broad faces of ferrite plates. *Acta Metall.* **27**, 365–76
Robertson J M 1929 The microstructure of rapidly cooled steel. *J. Iron Steel Inst.* **119**, 391–419
Sandvik B P J, Wayman C M 1983 Characteristics of lath martensite: II. The martensite–austenite interface. *Metall. Trans.* **14A**, 823–34
Shiflet G J, Aaronson H I 1990 Growth and overall transformation kinetics above the bay temperature in Fe–C–Mo alloys. *Metall. Trans.* **21A**, 1413–32
Spanos G, Fang H S, Aaronson H I 1990a A mechanism for the formation of lower bainite. *Metall. Trans.* **21A**, 1381–90
Spanos G, Fang H S, Sarma D S, Aaronson H I 1990b Influence of carbon concentration and reaction temperature upon bainite morphology in Fe-C-2 pct Mn alloys. *Metall. Trans.* **21A**, 1391–411
Troiano A R 1946 Discussion of kinetics of the decomposition of austenite. *Trans. AIME* **167**, 583–5
Wever F, Jellinghaus W 1932 Transformation kinetics of austenite: II. Dilatometry investigations of austenite decomposition. *Mitt. Kaiser-Wilhelm-Inst. Eisenforsch.* **14**, 85–9
Wever F, Lange H 1932 Transformation kinetics of austenite: I. Magnetic investigations of austenite decomposition. *Mitt. Kaiser-Wilhelm-Inst. Eisenforsch.* **14**, 71–83
Zener C 1946 Kinetics of the decomposition of austenite. *Trans. AIME* **167**, 550–83

W. T. Reynolds, Jr.
Virginia Polytechnic Institute and State University,
Blacksburg, Virginia, USA

Binary Oxide Ceramics: Al_2O_3, ZrO_2, Structure and Properties of

The structure of a material can, of course, be defined at multiple length scales ranging from subatomic to macroscopic. However, when considering what is usually meant by the term material properties, two levels of structure are of central interest—crystal structure (1–10 Å) and microstructure (1–100 μm). In addition, for oxide materials characterized by ionic interatomic bonding, defect structure (i.e., concentration and type of point defects) also frequently plays an important role.

When used in the engineering sense, "alumina" and "zirconia" really refer to nonmetallic alloys. Alumina is sometimes used as a single-phase pure material but zirconia never is, due to an inherent phase instability that takes place between normal processing and service temperatures. In both material systems, engineering control over structure is routinely accomplished through the introduction of intentional impurities. Furthermore, in most situations, the material is not allowed to adopt a fully equilibrated structure, but is used in a metastable, but kinetically persistent state. The consequence is that at the conclusion of processing, the material often reflects some of the characteristics of the raw material. Some specific examples are discussed in the following. As a result of this flexibility, both alumina and zirconia can be engineered to exhibit a wide range of properties.

Lastly, materials such as alumina and zirconia can be used as dense monolithic materials or a loose granular material. Often, but not always, in the latter situation, the material is treated as a consumable.

1. *Aluminum Oxide*

Alumina is the most widely utilized oxide ceramic. It is the only oxide ceramic widely used in single crystal form. Single crystal alumina, often incorrectly referred to as sapphire rather than corundum, is used for both its structural and optical properties. The only other significant uses of bulk single crystal oxides are artificial gemstones and laser hosts. The overwhelming majority of alumina products, however, use the polycrystalline form. The major markets for alumina-based materials on a weight basis are refractories (50%), abrasives (20%), whitewares and spark plugs (15%), and engineering ceramics (10%).

There is only one thermodynamically stable phase of aluminum oxide, *a*-alumina, which has the corundum structure, see Fig. 1. The crystal structure is often described as having O^{2-} anions in an approximately hexagonal close packed arrangement with Al^{3+} cations occupying two-thirds of the octahedral interstices. Although this idealized model of the crystal structure correctly predicts many of the properties, such as anisotropy of optical and elastic constants, it has been pointed out recently that the crystal structure deviates from this ideal, due primarily to electrostatic repulsion between aluminum ions, and that this loss of symmetry restricts plastic deformation at high temperature in important ways. The engineering properties of single-crystal *a*-alumina are given in Table 1. Of particular note are: the high possible use temperatures; the high values for hardness, stiffness, and strength; the moderate thermal expansion and density relative to structural metals; the low optical absorption of visible light; and the excellent electrical resistivity and corrosion resistance.

1.1 *Single Crystal Alumina*

Bulk single crystal alumina is exploited in a number of commercial applications. Historically, the first commercial uses of synthetic sapphires and rubies (a solid solution of chromium oxide in alumina) were as jewelry, abrasion resistant thread guides, components in the clockwork mechanism in watches, and draw plates for wire drawing (Belyaev 1980). In such applications the high hardness and excellent wear resistance, both at room and elevated temperatures, are the key properties. Optical properties dominate in another,

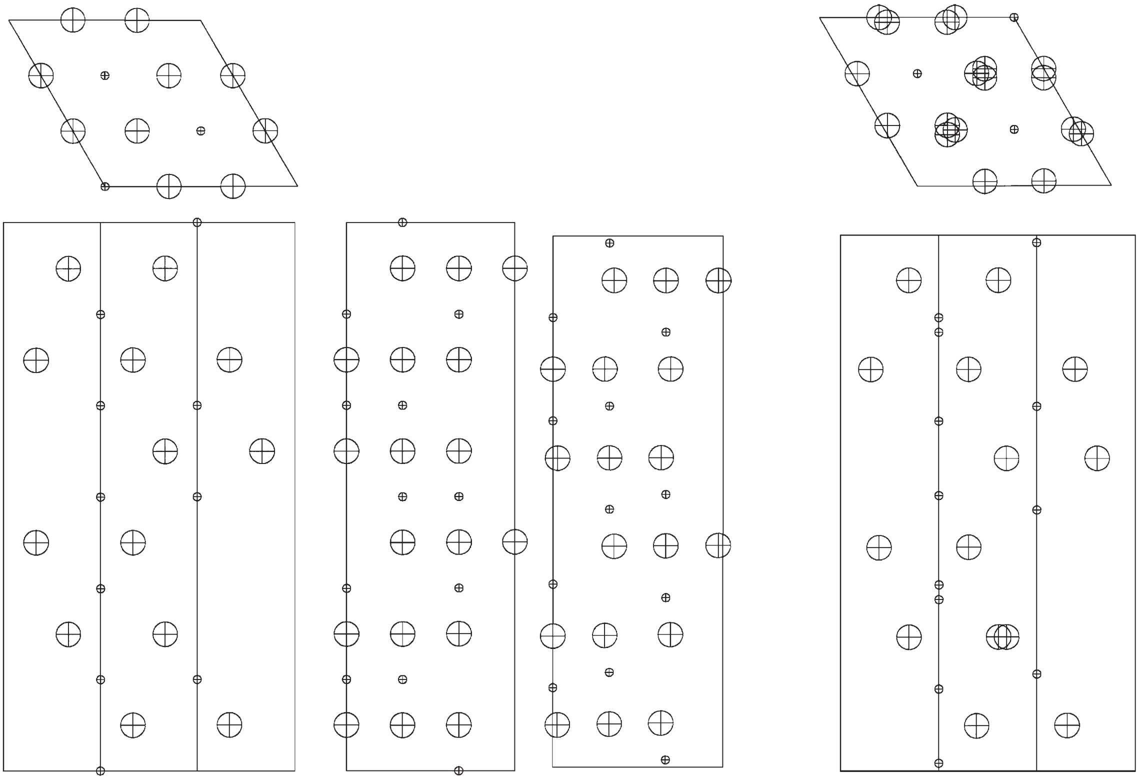


Figure 1
Top, front, and side views of a unit cell of the corundum structure showing the actual and commonly used idealized-structure on the right and left side of each view, respectively. The idealized description has hexagonal close-packed oxygen with two-thirds of the interstitial sites filled with aluminum. It conveys the correct general impression, but overestimates the symmetry displayed by Al_2O_3. The idealized cell has a c/a ratio of $\sqrt{8}$ and a volume equal to the true cell. Oxygen is depicted with the larger diameter.

perhaps more familiar, application of chromium-doped single-crystal alumina: as the active optical element for an important class of solid-state lasers. The crystal structure of the alumina host modifies the electron energy levels of the partially filled d-shell electrons in the chromium, which gives rise to the characteristic red color in contrast to the green color normally associated chromium ions.

The most flexible process for producing single crystal alumina from a melt is a modified Czochralski method, termed edge-defined film-fed growth (EFG), developed by LaBelle (1980). In this process the melt is shaped by extraction through a refractory metal tool as solidification takes place. Variants of the EFG process have led to the exploitation of the unique combination of properties offered by single crystal alumina in diverse applications. These include: substrates for silicon-on-sapphire integrated circuits, arc tubes for lighting applications, hollow fibers as optical waveguides for medical applications of lasers, and (one of the most important commercial applications) as abrasion resistant windows for supermarket laser scanners.

1.2 Alumina Ceramics and their Application

The vast majority of aluminum oxide is used as a polycrystalline solid. High density, high-alumina ceramics can be categorized as:

(i) those densified with the aid of a liquid by viscous flow; and

(ii) those where there is essentially no liquid phase and sintering is via the solid state.

The microstructures formed in such commercial aluminas are closely related to the macroscopic properties. Morrell (1985) subdivides commercial aluminas into six categories in terms of the wt.% alumina content. Compositions of greater than 99.7 wt.% alumina are referred to as solid-state sintered, and five distinct liquid-phase sintered composition ranges are

Table 1
Properties of single crystal alumina.

General	
Chemical formula	Al_2O_3 (aluminum oxide)
Names	corundum, sapphire, *a*-alumina
Crystal system	rhombohedral
Space group	$R\bar{3}c$
Thermal	
Melting point	2053 °C
Maximum useful temperature	~2000 °C
Specific heat	757 $Jkg^{-1}K^{-1}$ (25 °C), 1255 $Jkg^{-1}K^{-1}$ (1000 °C)
Thermal conductivity	40 $Wm^{-1}K^{-1}$ (25 °C), 10 $Wm^{-1}K^{-1}$ (1000 °C)
Thermal expansion coefficient (25–1000 °C)	$8.8 \times 10^{-6}\,K^{-1}$ (parallel to c-axis)
	$7.9 \times 10^{-6}\,K^{-1}$ (perpendicular to c-axis)
Physical/mechanical	
Density	3960 kgm^{-3} (25 °C)
Young's modulus (parallel to c-axis)	435 GPa (25 °C)
	386 GPa (1000 °C)
Shear Modulus	175 GPa (25 °C)
Poisson ratio	0.27–0.30 (orientation dependent)
Flexural strength (25 °C)	1035 MPa (parallel to c-axis)
	760 MPa (perpendicular to c-axis)
Compressive strength	~2 GPa (25 °C)
Hardness	9 Moh
	1900 Knoop (parallel to c-axis)
	2200 Knoop (perpendicular to c-axis)
Optical	
Refractive index	1.768 (ordinary ray, parallel to c-axis)
	1.760 (extraordinary ray)
Birefringence	0.008
Temperature coefficient of refractive index	$13 \times 10^{-6}\,K^{-1}$ (visible range)
Spectral emittance	0.1 (1600 °C)
Spectral absorption coefficient	0.1–0.2 cm^{-1} ($\lambda = 0.66$ mm, 1600 °C)
Electrical	
Volume resistivity	$10^{18}\,\Omega m^{-1}$ (25 °C)
	$10^{13}\,\Omega m^{-1}$ (500 °C)
	$10^{8}\,\Omega m^{-1}$ (1000 °C)
Dielectric strength	48 MVm^{-1} (1200 $Vmil^{-1}$)
Dielectric constant	11.5 (parallel to c-axis, $\upsilon = 10^3 - 10^9$ Hz, 25 °C)
	9.3 (perpendicular to c-axis)
Magnetic susceptibility	0.21×10^{-6} (parallel to c-axis)
	-0.25×10^{-6} (perpendicular to c-axis)
Chemical	
Acid resistance	insoluble in inorganic acids at room temperature
	attacked by boiling hydrofluoric acid
	slowly dissolved in molten salts and oxides at high temperature (>1000 °C)
Weathering resistance	unaffected by atmospheric exposure
Sea water resistance	unaffected by marine exposure
Biological resistance	unaffected by *in vivo* exposure
	nonthrombogenic and nonreactive with body fluids

identified: 88.9–99.7%, 96.5–99.0%, 94.5–96.5%, 86.0–94.5%, and 80.0–86.0%.

Solid-state sintering of polycrystalline alumina developed from a commercial need (see, for example, Bennison and Harmer 1990). In the late 1950s researchers at the General Electric Company laboratories in the USA required a translucent material that was resistant to alkali attack at high temperatures for use in envelopes in sodium-vapor discharge lamps, the sort now used extensively for street lighting. The primary problem was the inherent opacity in sintered material due to the presence of pores that efficiently scatter visible light. It was determined that the addition of 0.25 wt.% MgO enabled alumina to be sintered to a fine-grained, low porosity, translucent state after firing at 1900 °C in a hydrogen atmosphere (Coble 1961).

For many other applications, optical transmission of visible light is not a critical property and formulations are prepared that are lower cost and offer other attributes such as flatness and smoothness after firing and ease of metallization. Liquid-phase sintered, LPS, aluminas remain two-phase even after cooling to room temperature; the material that was liquid at the processing temperature, nearly always a silicate, converts to a glass or partially crystalline solid during cooling.

Pure α-Al_2O_3 is denser, harder, stiffer and more refractory than most silicate ceramics, so that increasing the proportion of second phase in an alumina ceramic tends in general to decrease the density, Young's modulus, strength, hardness, and refractoriness. However, fabricating products with the higher alumina contents is expensive, requiring pure starting materials and high firing temperatures. Intentional additions are made to alumina for a number of reasons including: lowering the firing temperature, allowing cheaper, less pure starting materials to be used, improving rheology in shape forming, and modifying the properties of the product (Morrell 1987). A broad range of materials is commercially available with a concomitant broad range of properties. Trends in data given in the following sections are meant as a general guide. Other microstructural variables such as porosity, grain size and second phase composition may also have important effects on properties such as strength and thermal conductivity.

Solid-state sintered alumina is produced from high purity powders, which densify to give single-phase ceramics with uniform grain size. Products fired in air at 1600–1700 °C may contain some residual porosity. The reason for the persistence of $\sim$5% porosity is simply that in the last stages of sintering all of the pores are isolated within the oxide matrix. Therefore, further shrinkage of the pore requires the gas within the pore to dissolve in the oxide and diffuse to the external surface. Nitrogen is not soluble in alumina at the sintering temperature and therefore the pores cannot be completely removed.

Coarse-grained translucent alumina, used for sodium vapor lamp envelopes, is fired at 1700–1800 °C in hydrogen, which has a relatively large solubility and high diffusivity, to improve the rate of elimination of porosity. It is fired at high temperatures to increase the average grain size. Since alumina has an index of refraction that is anisotropic, some light scattering occurs at each grain boundary. The large grain size reduces the linear density of grain boundaries and thereby increases the total transmission for visible light. When 99.8% Al_2O_3 is used in very high temperature refractory applications, for example as crucibles, MgO cannot be used since it is prone to evaporation at high temperature. As a result, aluminas used in these applications typically have coarse-grained microstructures. These "recrystallized" alumina refractories are relatively weak as a result of the large mean grain size. When small amounts of MgO, even as low as 0.5%, are added to control grain growth, the resultant LPS alumina is much stronger and can be used in applications requiring high-temperature insulating ability, i.e., in kiln furniture and thermocouple insulation and for small section and thin walled components, e.g., rods or tubes.

Many engineering alumina ceramics contain 99–99.7% alumina and so have a significant glass content. Grain sizes in the range 2–25 μm are typical and the products can be particularly strong. The fine grain sized 99% aluminas may be used in demanding applications such as hip prostheses, whereas coarser grained materials are preferred for electrical insulation.

Aluminas with 94.5–99% Al_2O_3 have a large proportion of grain boundary glass that must be of carefully controlled composition to confer the required densification behavior and final state properties. The grain boundary glass is usually an aluminosilicate containing additional oxide such as CaO or MgO. Microstructures in aluminas where the glass is used simply as a densification aid show a uniform distribution of alumina crystals completely separated by glass. In other aluminas that are fired to higher temperatures, some recrystallization of the alumina may occur to give an interconnected network. In these materials pores are usually located at the interface between the alumina and the glass. Typical properties are given in Table 2.

Although these aluminas cannot be used in applications requiring very high temperature stability, they can be tailored for use in many electrical applications. The second phases present depend in general on the composition used, the firing temperature, and the cooling rate (more glass being present with faster cooling). The ease of crystallization was observed to be strongly dependent on the ratio of MgO to CaO in the intergranular phase (Powell-Dogan and Heuer 1990a, 1990b, 1990c). The mechanical properties of glass-containing alumina ceramics depend on the thermal expansion coefficient of the glass and the extent of crystallization as the volume change associated with crystallization causes microcracking, which degrades the fracture strength.

The presence of a grain boundary glass is not, however, necessarily deleterious. For example, the metallization of a substrate is often necessary to allow brazing or soldering and to provide a conducting surface. Procedures such as the molymanganese process (Kohl 1967) or thick-film circuitry (Harris and

Table 2
Properties of sintered high-purity and debased alumina.

Al_2O_3 wt. %	>99.9	99.9	99.7–99.9	99.0–99.7	96.5–99.0	94.5–96.5	86.0–94.5	80.0–86.0
Density	3.98	3.6–3.8	3.6–3.8	3.7–3.8	3.9–3.96	3.7–3.9	3.4–3.7	3.3–3.4
Hardness (Gpa) (Vickers 500 g)	19.3	16.3	15–16	15–16	12.8–15	12–15.6	9.7–12	
Fracture toughness, K_{IC}(RT, MPa $m^{1/2}$)	2.8–4.5			5.6–6				
Young's modulus (GPa)	366–410	300–380	300–380	330–400	300–380	300	250–300	200–240
Bend strength (MPa)	550–600	160–300	245–412	550	230–350	310–330	250–330	200–300
Thermal expansion coefficient ($\times 10^{-6}\,°C^{-1}$)	6.5–8.9	5.4–8.4	5.4–8.4	6.4–8.2	8–8.1	7.6–8	7–7.6	
Thermal conductivity ($W\ m^{-1}K^{-1}$)	38.9	28–30	30	30.4	24–26	20–24	15–20	
Firing range (°C)	>1800	>1750	>1750	>1700		1520–1600	1440–1600	

Lall 1991) rely on interactions during firing between the powder particles, which are painted on, and preexisting intergranular glass. For example, in thick-film metallization, a mixture of desired metals or conductors and a low temperature melting glass frit is painted on the desired region and the substrate. During subsequent firing, the metal particles fuse together to provide the conduction path while the glass particles fuse and begin to dissolve the more refractory glass in the ceramic substrate. The result is a strong bond with graded properties.

Aluminas with 80–94.5% Al_2O_3 are generally used as electrical insulators, low-temperature mechanical components, or refractories. The glass-bonded insulators and mechanical components are fired at greater than 1500 °C and may suffer viscous flow of the glass at comparatively low temperatures. If nonequiaxed alumina particles are used, the fluid nature of the glass during processing may allow their reorientation, leading to anisotropic grain structures and properties.

1.3 Other Applications of Alumina

In addition to its use as a densified ceramic, alumina also is used in other forms such as powders or fibers. In some applications, they are used as a freely flowing granular material, and in others they are partially sintered, usually just enough to develop some strength.

One such application employing *a*-alumina is as an abrasive. Both grinding wheels and coated abrasive paper (emery) are used in addition to free abrasive. One situation in which alumina grinding wheels are preferred to, say, silicon carbide is when working materials that have a high sensitivity to heat. The high thermal conductivity and high heat capacity coupled with the friable nature of the wheel allow most of the heat produced during to grinding to be carried away from the work-piece.

Despite the fact that there is only one equilibrium phase, many processes such as the oxidation of aluminum metal and the thermal decomposition of gibbsite or vapor-deposited amorphous alumina thin films involve the formation of intermediate metastable alumina phases. These so-called transition phases are denoted as γ, χ, η, ι, ε, δ, ϕ, and κ. The most important commercially is γ, which is used both as a polishing compound and as a catalyst substrate because of its fine particle size and acicular, or needlelike, shape.

Aluminosilicate ceramic fibers based on alumina have been commercially available for many years. The largest use for high-alumina aluminosilicate refractory fibers (containing less than 28 wt.% silica) is as high-temperature low-thermal-mass furnace insulation. The noncontinuous or wool-like fibers are used as loose wool, blankets, felts, paper, and board. The loose wool is made by melt spinning or air jet blowing of kaolin-based materials. Fibers with 52% alumina can be used as insulation up to 1250 °C and those containing 65% Al_2O_3 can be used up to 1450 °C. Materials with higher percentages of alumina are capable of extended service at temperatures in excess of 1700 °C. The advent of such low thermal-mass furnace insulation played an important role in re-engineering furnaces to increase energy efficiency during the last two decades of the twentieth century.

2. *Zirconium Dioxide*

Zirconia is also a refractory oxide and shares with alumina the ability to serve some high-temperature applications. However, zirconia exists as three polymorphs (see Fig. 2), depending on temperature, and this richness offers the possibility of greater engineering of properties. Point defect chemistry, also, is more readily engineered in zirconia-based systems. The traditional applications of ZrO_2 and ZrO_2-containing

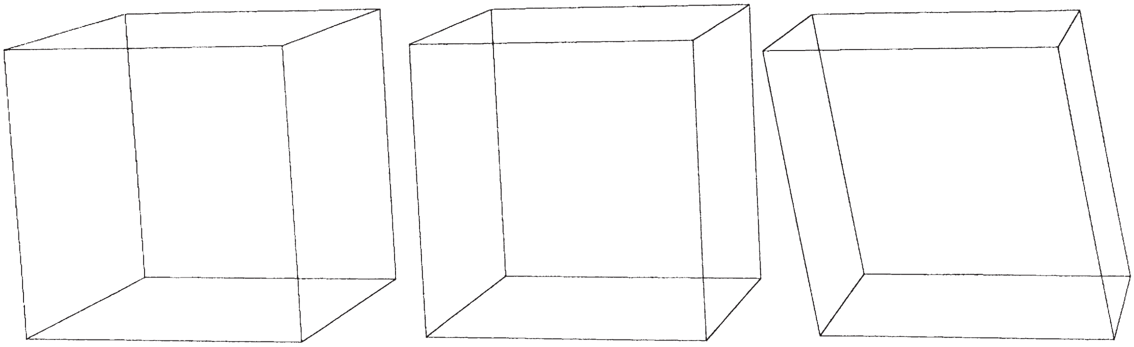


Figure 2

Perspective drawings of the shapes of the unit cells associated with the polymorphs of zirconia, drawn to scale. From left to right, the cells correspond to cubic, tetragonal, and monoclinic polymorphs. The cubic to tetragonal cells are very similar, but the monoclinic cell differs significantly in both shape and volume ($V_m > V_t$). Cell parameters used to create the figure are: for cubic $a = b = c = 0.5113$ nm and $\alpha = \beta = \gamma = 90°$; for tetragonal $a = b = 0.5082$ nm, $c = 0.5185$ nm, $\alpha = \beta = \gamma = 90°$; and for monoclinic $a = 0.5147$ nm, $b = 0.5206$, $c = 5.135$, $\alpha = \gamma = 90°$ and $\beta = 99.23°$. (Precise values depend on both stabilizer type and content as well as temperature.)

materials are foundry sands and flours, refractories, ceramic and paint pigments, and abrasives. These applications account for most of the tonnage used. However, the thermomechanical and electrical properties of zirconia-based ceramics have led to a wide range of advanced and engineering ceramic applications. Tough, wear resistant and refractory, ZrO_2 is being developed for applications such as extrusion dies, machinery wear parts, and piston caps. Composites containing ZrO_2 as a toughening agent such as ZTA (zirconia toughening alumina) also show promise in applications such as cutting tools. Ionically-conducting ZrO_2 can be used as a solid electrolyte in oxygen sensors, fuel cells, and furnace elements.

2.1 Crystallography of Zirconia

Zirconia occurs in three well-established polymorphs: monoclinic (m), cubic (c), and tetragonal (t). Table 3 lists some typical properties of single-crystal zirconia. In pure ZrO_2 the monoclinic phase is stable up to 1170 °C; above this temperature it transforms to tetragonal symmetry and then to cubic symmetry at 2370 °C before melting at 2680 °C. The transformation from monoclinic to tetragonal exhibits a hysteresis; the transformation on cooling occurring 100 °C below 1170 °C. The transformation to monoclinic is martensitic, and during cooling results in a volume increase of the order of 3–4%. This volume change is sufficient to exceed the elastic limit of the ZrO_2 grains and cause cracking.

2.2 Stabilized Zirconias

Most applications of zirconia therefore require that the structure be fully- or partially-stabilized through alloying with alkaline earth or rare earth oxides. The term "stabilized" refers to a kinetic stabilization of a solid solution in the cubic polymorph to room temperature. Full stabilization refers to compositions that exhibit single-phase behavior from absolute zero to the solidus. Even with such alloys, thermal shock resistance is still an important issue. This is due to their high thermal expansion coefficient and low thermal conductivity. The low thermal conductivity tends to cause steep thermal gradients during heating or cooling and the high thermal-expansion coefficient results in large thermal strains or high stresses. The discovery that thermal shock resistance was improved by only adding enough of the alloying element to partially stabilize the cubic phase led to the use of the partially stabilized zirconia (PSZ) as a refractory and its later development as an engineering ceramic due to its high toughness arising from transformation toughening, a development pioneered by Garvie *et al.* (1975). High toughness is achieved when a microstructure is developed that contain metastable tetragonal particles constrained in a cubic matrix. The precipitates transform to monoclinic symmetry when a propagating crack relieves the constraint. The volume change and shear strain associated with the martensitic reaction oppose the opening of the crack, so increasing the resistance of the ceramic to crack propagation, i.e., increasing its toughness (Green *et al.* 1989, Evans 1984, Claussen 1984).

In addition to the most commonly used stabilizers (the oxides of calcium, magnesium, cerium, and yttrium), virtually all the rare earth elements form solid solutions with zirconia. In general, where the zirconium ion is in eight fold coordination (i.e., tetragonal or cubic), ions will stabilize the zirconia phase provided that the ionic radius is within approximately 40% of that of Zr^{4+}. In addition, many alloy

Table 3
Properties of single crystal zirconia and zirconia-based alloys.

General	
Chemical formula	ZrO_2 (zirconium oxide)
Names	baddeleyite
Crystal systems and space groups	monoclinic, P21/c (up to 1170 °C)
	tetragonal, P42/nmc (1170–2370 °C)
	cubic, Fm3m (2370–2680 °C)
Thermal	
Melting point	2680 °C
Maximum useful temperature	~2400 °C
Thermal conductivity	$1.5\,Wm^{-1}K^{-1}$ (25 °C), $2\,Wm^{-1}K^{-1}$ (1000 °C)
Thermal expansion coefficient (25–1000 °C)	
Monoclinic	$7 \times 10^{-6}\,°C^{-1}$ (parallel to a-axis)
	$2 \times 10^{-6}\,°C^{-1}$ (parallel to b-axis)
	$13 \times 10^{-6}\,°C^{-1}$ (parallel to c-axis)
Tetragonal	$9 \times 10^{-6}\,°C^{-1}$ (parallel to a-axis)
	$12 \times 10^{-6}\,°C^{-1}$ (parallel to c-axis)
Cubic	$7.5–13 \times 10^{-6}\,°C^{-1}$ (dependent on stabilizer type and amount)
Physical/mechanical density	
Monoclinic	$5830\,kgm^{-3}$
Tetragonal	$5860\,kgm^{-3}$
Cubic	$5640\,kgm^{-3}$ (stabilizer type and amount can alter density)
Optical (stabilized cubic form)	
Refractive index	2.15–2.18
Birefringence	none
Dispersion	0.058–0.060
Spectral emittance	0.2 (visible wavelengths)

additions promote anion ordering (the cations remain disordered) rendering a more stable lattice.

The processing of partially stabilized zirconias to tailor the microstructure is, similar to metals such as steel, primarily one of controlling the heat treatment in order to effect a particular solid state transformation sequence.

For example, a typical composition for a magnesia-partially stabilized zirconia, Mg-PSZ, may be around 8 mol.% MgO. The first step in processing involves a solution heat treatment in the cubic single phase field (2–4 h at about 1800 °C depending on composition), followed by a rapid cool. This quench is too rapid to allow the equilibrium amount of tetragonal phase to precipitate out, but it does promote homogeneous nucleation of very fine tetragonal precipitates. The maximum cooling rate must, however, be limited to avoid thermal shock. Reheating to 1400 °C and holding isothermally (aging) leads to coarsening of the tetragonal particles by rejection of MgO into the cubic matrix. If overaging is avoided, the resultant microstructure is one of finely divided precipitates in a cubic matrix. Small tetragonal particles, 0.2 μm, may be metastably retained upon cooling due to the presence of the cubic matrix. Coarser particles spontaneously transform to monoclinic symmetry during cooling.

If the tetragonal precipitates are above a critical size they will transform to monoclinic symmetry either spontaneously or as a result of applied stress. It should be noted that this critical size depends on many factors including the degree of constraint (whether particles are in a bulk matrix or powder form), temperature, and composition.

Commercial Mg-PSZ often undergoes a further subeutectoid aging treatment at 1100 °C to improve the room-temperature properties. The microstructure of commercial PSZ is rarely optimized; more usually it is that produced by a furnace cool after sintering in the single cubic phase field, or possibly a rapid cool from the sintering temperature to an isothermal hold temperature. This heat treatment schedule allows more heterogeneous nucleation of tetragonal particles to take place. Consequently, a thicker grain boundary tetragonal film is produced which spontaneously transforms to monoclinic on cooling. In addition, heterogeneously nucleated precipitates form within the grains. Such precipitates grow rapidly during a subsequent isothermal hold at 1400 °C. The precipitates thus formed, are also large enough to transform to monoclinic on cooling, thereby reducing the total amount of residual metastable tetragonal phase available for transformation toughening.

There are strong analogies between Y-PSZ, Ca-PSZ, and Mg-PSZ; similar heat treatments are used to develop optimal microstructures in all three. Representative mechanical and thermal properties for commercial PSZs are given in Table 4. While a useful guide, the data should be treated with caution since the particular values obtained depend on the test method, especially for K_{IC}. Further, these general values are of course, affected by material variables in the ceramic microstructure (stabilizer content, grain size, etc.) and external variables such as atmosphere and temperature.

2.3 *Fully Tetragonal Zirconias*

In the case of Y_2O_3–ZrO_2, however, there is extensive solubility for Y_2O_3 in the terminal tetragonal solid solution. Up to approximately 2.5 mol.% Y_2O_3 will be taken into solid solution, which in conjunction with the low eutectoid temperature, allows a fully tetragonal ceramic to be obtained (so-called tetragonal zirconia polycrystals or TZP) first reported by Rieth *et al* (1976). Fine grain sizes are obtained by using ultrafine powders and sintering in the range 1400–1500 °C where the coarsening rate can be controlled. Similar to the critical precipitate size in PSZ materials, TZP ceramics exhibit a critical grain size (about 0.3 μm), above which spontaneous transformation occurs leading to low strength and toughness. The critical size depends on the composition (being about 0.2 μm for 2 mol.% Y_2O_3 and 1.0 μm for 3 mol.% Y_2O_3) and the degree of mechanical constraint.

TZPs have a significant advantage over Mg-PSZs because sintering can be carried out at comparatively low temperatures (1400–1500 °C compared to 1800 °C) bringing the production of TZPs within the scope of existing furnaces in ceramic manufacturing plants. The bulk of commercial TZPs contain 2–3 mol.% Y_2O_3 and mainly consist of fine equiaxed tetragonal grains of a diameter typically in the range 0.2–2 μm.

TZPs also can be formed with cerium as the stabilizer and there are many similarities between the Ce and Y-TZPs although a fully tetragonal structure can be obtained in Ce-TZP for CeO_2 additions between 12 and 20 mol.%, representing a much wider range than with Y-TZPs (Tsukuma and Shimada 1985).

For an equivalent fracture toughness, the grain size is larger in Ce-TZP compared to Y-TZPs. For example, a K_{IC} of 12 MPa $m^{1/2}$ could be achieved in a Y-TZP with a grain size of 2 μm compared to 8 μm in a Ce-TZP (Tsukuma and Shimada 1985). Very high toughness values can be achieved in these materials, with nominal values of up to 30 MPa $m^{1/2}$ being reported, although the absolute value depends strongly on the test method used.

Typical mechanical and thermal properties for commercial TZPs are given in Table 5. The exact yttria content in Y-TZP plays an important role in the transformability of the tetragonal phase and therefore the toughness. Toughness is also a strong function of grain size.

Table 4
Properties of partially-stabilized zirconia-based alloys.

	Mg-PSZ	Ca-PSZ	Y-PSZ
Stabilizer (wt. %)	2.5–3.5	3–4.5	5–12.5
Hardness (GPa)	14.4	17.1	13.6
Fracture toughness, K_{IC}(RT, MPa $m^{1/2}$)	7–15	6–9	6
Young's modulus (GPa)	200	200–217	210–238
Bend strength (MPa)	430–720	400–690	650–1400
Thermal expansion coefficient ($\times 10^{-6}$ °C^{-1})	9.2	9.2	10.2
Thermal conductivity(W $m^{-1}K^{-1}$)	1–2	1–2	1–2

Table 5
Properties of tetragonal zirconia polycrystals.

	Y-ZP	Ce-TZP
Stabilizer (mol.%)	2–3	12–15
Hardness (GPa)	10–12	7–10
Fracture toughness, K_{IC} (RT, MPa$m^{1/2}$)	6–15	6–30
Young's modulus (GPa)	140–200	140–200
Bend strength (MPa)	800–1300	500–800
Thermal expansion coefficient ($\times 10^{-6}$ °C^{-1})	9.6–10.4	
Thermal conductivity(W $m^{-1}K^{-1}$)	2–3.3	

2.4 Applications of Zirconia

Zirconia ceramics have the highest toughness of any ceramic, which, combined with high strength, hardness and chemical resistance, should allow their application in harsh environments under severe loading conditions. Wear-resistant applications of Mg-PSZ as drawing dies, bearings, seals, and bone-replacement devices (mostly ball and socket joints) have been developed. The low thermal conductivity of ZrO_2 can be used to advantage in automotive engines in piston crowns, head face plates, and piston liners so that heat loss from the combustion chamber can be reduced and flame temperature increased resulting in higher efficiency. Wear resistant applications in engines include those in the valve train as cams, cam followers, tappets, and exhaust valves. The thermal expansion coefficient of zirconia closely matches that of cast iron so that these materials can be joined (typically by an active substrate process with a titanium-based bond) to give relatively inexpensive automotive components.

PSZ refractory crucibles, shaped by slip casting or isostatic pressing are used in vacuum induction or air melting of refractory metals such as cobalt-base alloys or precious metals such at platinum, palladium, or rhodium. The relatively high cost of ZrO_2 compared to other refractories limits its use to special applications where it has particularly desirable properties. PSZ is comparatively stable to both acidic and basic slags and molten steel. This amphoteric behavior, combined with its superior erosion and thermal shock resistance, enables its use for submerged entry tundish nozzles in continuous casting of steel. PSZ inserts are also used in alumina–graphite sliding and rotary gate valves when high oxygen or calcium-bearing steels are used and increased corrosion/erosion resistance is required. Zirconias are also used in a number of cutting applications for difficult materials such as glass fibers, magnetic tape, plastic film and paper items such as cigarette filters (Stevens 1986).

2.5 Ionic Conductivity of Stabilized Zirconia

The same stabilizing process that permits the engineering of mechanical properties also changes the point defect chemistry producing a material that is an excellent ionic conductor, e.g., yttria-stabilized zirconia system. One way of describing the process is as a dissolution process, according to the following chemical reaction, written in Kroger–Vink notation (Kroger 1974) as:

$$Y_2O_3 \overset{ZrO_2}{\rightarrow} 2Y'_{Zr} + 3O_O^x + V_O^{\bullet\bullet}$$

That is, because the yttria is dissolving to form a solid solution with a well defined crystal structure it must form anion and cation sites in the correct ratio; in this case 2:1. Since the stabilizer does not supply enough anions, relative to the number of cations it supplies, its presence produces oxygen vacancies (indicated above as V_O). Relatively high concentrations of vacancies can be readily produced in stabilized zirconias, with up to 10% or 15% of the oxygen sites vacant. Another aspect of the zirconia crystal structure is that the mobility of oxygen vacancies is relatively high.

The consequence of this is that there is a relatively wide temperature regime, approximately 500–1000 °C, over which stabilized zirconias exhibit ionic conductivity, but essentially no electronic conductivity. That is, all of the electrical current that flows through the solid is the result of the movement of oxygen ions not electrons.

This property combined with the high refractoriness and chemical durability of the material leads to its application in a wide array of devices. PSZ is used as oxygen sensors, both in automobile combustion feedback systems and industrial process control. One emerging application likely to grow in importance is the use of PSZ as the solid electrolyte in fuel cell applications. Finally, ionic conductors such as PSZ can be used as selective membranes for selective oxygen enrichment systems used in medical applications.

2.6 Zirconia-toughened Ceramics

Zirconia has been added to a variety of other oxide matrices such as mullite, spinel, cordierite, zircon, and MgO in order to improve their toughness. The first system was zirconia-toughened alumina (ZTA), developed simultaneously with the PSZ systems (Claussen 1976). This material demonstrated that the inclusion of unstabilized zirconia can lead to retention of t-ZrO_2 in the sintered product if the particle size is small enough. The toughening mechanism primarily involves transformation and microcrack formation although dispersion strengthening is also active in this system. Less than 20 vol.% unstabilized zirconia particles are typically used although subsequent work has attempted to increase the ZrO_2 content by partial stabilization of the zirconia with Y_2O_3, CeO_2, or TiO_2 (Lange 1982). Unstabilized ZTAs can have strengths of 1200 MPa and toughnesses of 16 $MPam^{1/2}$ with about 15 vol.% ZrO_2 compared to values of 600 MPa and 4 MPa $m^{1/2}$ respectively for typical dense alumina.

ZTA was first used as a toughened abrasive for industrial grinding wheels where large improvements in grinding efficiency were detected over conventional materials. Other applications are found in metal cutting tools and engine components.

3. ***Summary***

Al_2O_3 and ZrO_2 are both "workhorse" ceramics used in a wide variety of applications that exploit different properties: mechanical, optical, thermal, and electrical. Both are high-temperature materials used in refractory applications. Both are hard and wear-resistant and are used in applications involving sliding and wear. Alumina has only stable phase, whereas zirconia has three polymorphs. This leads to different strategies for tailoring the microstructure in each case.

With ZrO_2, proper chemistry and careful microstructural control allow phase transformations to be exploited to yield ceramics with outstanding strength and toughness. Zirconia is being applied to uses previously thought to be outside the realm of ceramics, notably in engine components. Furthermore, its unique electrical properties lead to a wide variety of applications in sensors. Microstructural tailoring of alumina involves the inclusion of additives to regularize the grain structure and the inclusion of a second phase silicate.

Bibliography

Belyaev L M 1980 In: Belyaev L M (ed.) *Ruby and Sapphire*. Amerind, New Delhi, India, pp. vi–xiv

Bennison S J, Harmer M P 1990 A history of the role of MgO in the sintering of α-$A1_2O_3$. In: Handwerker C A, Blendell J E (eds.) *Ceramic Transactions*, Vol. 7: *Sintering of Advanced Ceramics*, American Ceramic Society, Westerville, Ohio, USA, pp. 13–49

Claussen N 1976 Fracture toughness of $A1_2O_3$ with an unstabilized ZrO_2 dispersed phase. *J. Am. Ceram. Soc.* **59**, 49–51

Claussen N 1984 Microstructural design of zirconia-toughened ceramics ZTC. Claussen N, Ruhle M, Heuer A H (eds.) *Advances in Ceramics*, Vol. 12: *Science and Technology of Zirconia II*, American Ceramic Society, Westerville, Ohio, pp. 325–51

Coble R L 1961 Sintering crystalline solidas. II Experimental test of diffusion models in powder compacts. *J. Appl. Phys.* **32**, 793–9

Evans A G 1984 Toughening mechanisms in zirconia alloys. In: Claussen N, Ruhle M, Heuer A H (eds.) *Advances in Ceramics,* Vol. 12: *Science and Technology of Zirconia II. Advances in Ceramics*. American Ceramic Society, Westerville, Ohio, pp. 193–212

Garvie R C, Hannink R H, Pascoe R T 1975 Ceramic steel? *Nature* **248**, 703

Green D J, Hannink R H J, Swain M V 1989 *Transformation Toughening of Ceramics*. CRC Press, New York

Harris D H, Lall P 1991 Printed wiring board and fabrication. In: Pecht M (ed.) *Handbook of Electronic Package Design*. Marcel-Dekker, New York, pp. 101–52

Kohl W H 1967 *Handbook of Materials and Techniques for Vacuum Devices*. Reinhold, New York

Kroger F A 1974 In: *The Chemistry of Imperfect Crystals*, Vol. 2 North-Holland, New York, p.14

LaBelle H E 1980 EFG, invention and application to sapphire growth. *J. Cryst. Growth* **50**, 8–17

Lange F F 1982 Transformation toughening part 5: effect of temperature and alloy on fracture toughness. *J. Mater. Sci.* **17**, 255

Morrell R 1985 *Handbook of Properties of Technical and Engineering Ceramics*, Part 1: *An Introduction for the Engineer and Designer*. HMSO, London

Morrell R 1987 *Handbook of Properties of Technical and Engineering Ceramics*, Part 2: *Data Reviews*. HMSO, London, Sect. 1

Powell-Dogan C A, Heuer A H 1990a Microstructure of 96% alumina ceramics: I, characterization of the as-sintered materials. *J. Am. Ceram. Soc.* **73**, 3670–6

Powell-Dogan C A, Heuer A H 1990b Microstructure of 96% alumina ceramics: II, crystallization of high magnesia boundary glasses. *J. Am. Ceram. Soc.* **73**, 3677–83

Powell-Dogan C A, Heuer A H 1990c Microstructure of 96% alumina ceramics: III, crystallization of high calcia boundary glasses. *J. Am. Ceram. Soc.* **73**, 3684–91

Rieth P H, Reed J S, Naumann A W 1976 *Bull. Am. Ceram. Soc.* **55**, 717–21, 727

Stevens R 1986 *An Introduction to Zirconia*. Magnesium Elektron, Twickenham, UK

Tsukuma K, Shimada M 1985 Strength, fracture toughness and Vickers hardness of CeO_2-stabilized tetragonal ZrO_2 polycrystals (Ce-TZP). *J. Mater. Sci.* **20**, 1178–84

J. D. Cawley
Case Western Reserve University, Cleveland,
Ohio, USA

Block Copolymer Phase Behavior

This article provides an overview of the microphase behavior of block copolymers and block copolymers blended with solvent or homopolymer, and the thermodynamics controlling them. The block copolymers can be amorphous or semicrystalline. The article covers microphase separation, and some aspects of the interplay of microphase and macrophase separation. Specific topics include the microphase diagrams of conformationally symmetric and asymmetric copolymers, the microphase diagrams of copolymer/solvent blends, induced microphase formation, micelle formation, and the scaling behavior of domain sizes in amorphous and semicrystallizable copolymers.

1. Introduction

Homopolymer blends normally phase separate. This *macro*phase separation occurs because of the unfavorable interactions between unlike monomers. *Block copolymers* are composed of distinct chemical species which are chemically bonded together within the individual molecules, such as poly(styrene)-*b*-poly(isoprene) (PS-*b*-PI). Copolymers cannot phase separate into macroscopic regions of different species because of their chemical bonds. Instead, they can *micro*phase separate, forming domains whose dimensions are limited by the sizes of the constituent blocks. In copolymer/homopolymer or copolymer/solvent mixtures, both microphase and macrophase separation can occur.

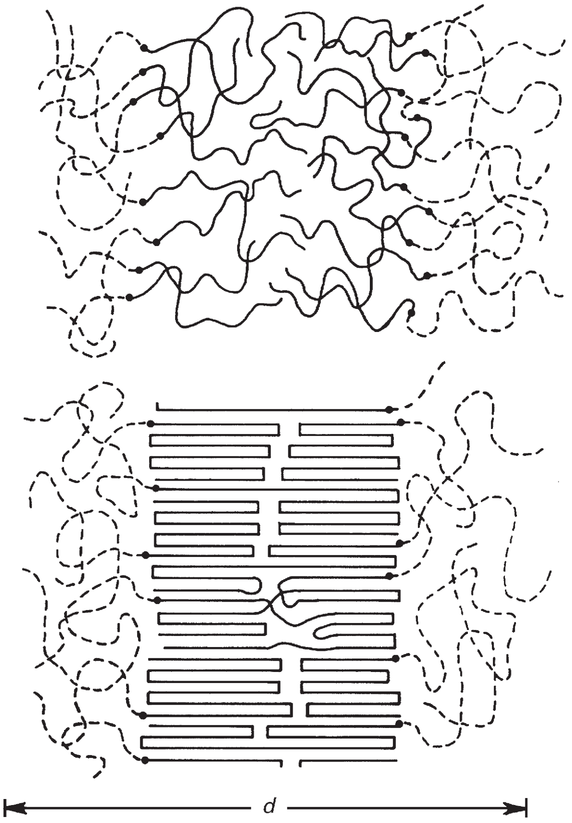


Figure 1
Schematic diagram of layered structures of microphase-separated, diblock copolymers. The upper and lower panels illustrate amorphous and semicrystallizable copolymers, respectively. Both panels correspond to strong segregation, with each block segregated into its respective subdomains, and the copolymer joints localized within narrow interphase regions between the subdomains. The layer repeat distance is *d*. As indicated near the center of the lower panel, the crystallizable block is generally semicrystalline, and the chain folding

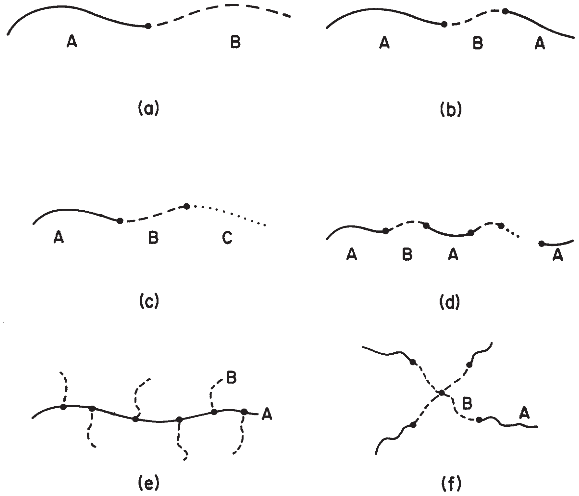


Figure 2
Schematic diagram of common copolymer architectures: (a) diblock A–B copolymers; (b) linear triblock A–B–A copolymers; (c) linear triblock A–B–C copolymers; (d) linear multiblock A–B–A–B– … copolymers, which can terminate with either an A or a B block; (e) graft copolymer, illustrated with B side chains grafted to an A

The transition from a disordered phase (D) to a microphase is generally referred to as the microphase separation transition (MST). Various domain structures occur, which can form either ordered or disordered arrays. In most cases, the material within each domain is amorphous, but in some systems one or more of the blocks can crystallize. The simplest structure consists of alternating layers, or lamellas. It is shown schematically for both amorphous and semicrystallizable copolymers in Fig. 1.

PS-*b*-PI is a diblock copolymer, generically labeled A-*b*-B, or simply A–B. There are also triblock copolymers, of the form A–B–A or A–B–C, linear multiblocks, e.g., A–B–A–B–, and more complicated architectures such as starblocks or graft copolymers. These are illustrated in Fig. 2.

The goal here is to introduce the current experimental and theoretical pictures of the equilibrium phases and domain sizes of systems containing copolymers. An important point is that the equilibrium microphase diagrams of all copolymers are very similar in structure, with only small differences that depend on molecular architecture and species. The theoretical understanding is presented in terms of numerical self-consistent field theory. This theory is also compared briefly with earlier theories introduced to treat weak and strong segregation regimes, and with experiments. This latter comparison is complicated by the fact that nonequilibrium effects are very common in experimental systems.

2. *Amorphous Block Copolymers*

2.1 *Microphase Diagrams*

The existence of the MST can be easily understood from simple thermodynamics. In the disordered phase of a system of neat A-*b*-B copolymers, each monomer interacts with like and unlike monomers. The effective interaction of two unlike monomers is normally repulsive, and these interactions tend to induce phase separation. The number of repulsive contacts per molecule, and hence the repulsive energy per molecule, increase with the molecular weight of the polymer. The phase separation is accompanied by a decrease in entropy, but the associated change in free energy per molecule is essentially independent of molecular weight. This entropic contribution increases with temperature, and the net result is that the disordered phase is stable for sufficiently high temperature and low molecular weight. However, the system microphase separates for decreased temperature and/or increased molecular weight. Leibler (1980) predicted that the MST for perfectly symmetric copolymers occurs at

$$\chi Z_c = 10.5 \qquad (1)$$

where χ is the Flory-Huggins A–B interaction parameter, Z_c is the polymer degree of polymerization, and the layer thickness at the MST is independent of χ and scales as

$$d \propto Z_c^{1/2} \qquad (2)$$

The system is disordered for $\chi Z_c < 10.5$, and ordered for $\chi Z_c > 10.5$. As shown in Fig. 3, the MST moves to higher χZ_c for compositionally asymmetric copolymers since, for a given Z_c, the number of unlike contacts per molecule varies as $f_A f_B$, where f_A and $f_B = 1 - f_A$ are the volume fractions of the A and B blocks.

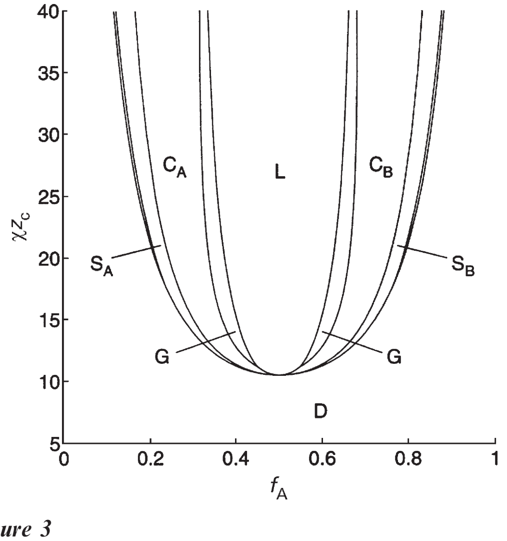


Figure 3
Microphase diagram for conformationally symmetric, diblock copolymers calculated by Matsen and Bates (1997) using numerical self-consistent field theory. The MST separates the disordered phase (D) from the microphases. The labeled microphases are lamellas (L), gyroid (G), A- and B-centered cylinders on a hexagonal lattice (C_A and C_B, respectively), and A- and B-centered spheres on a b.c.c. lattice (S_A and S_B, respectively). The narrow regions adjacent to the MST are the close-packed spheres (courtesy of Dr M. Matsen).

Mean field theories, including Leibler's, predict the MST to be second order at $f_A = 0.5$, and first order otherwise. Hence, the point $f_A = 0.5$ and $\chi Z_c = 10.5$ is a mean field critical point. Microphase-separated systems in the vicinity of this point are weakly segregated, with significant intermixing of the different components.

There are a number of microphases that can form (Bates 1991, Matsen and Bates 1997, Bates and Fredrickson 1999). The simplest one is the alternating layer structure of Fig. 1, in which each layer is relatively rich in one of the components, e.g., A–B–A–B– .... We will denote this phase simply as L. Another consists of parallel cylinders, each rich in one of the components, arranged on a hexagonal lattice. We denote these by C, or by C_A and C_B where the subscript labels the primary component of the interiors of the cylinders. A third consists of A or B rich spheres, which can be arranged on a body-centered cubic (b.c.c.-S) or a close packed (c.p.-S) lattice, or can be disordered. The L, C, and b.c.c.-S phases have traditionally been labeled as the "classical" phases. The best understood "exotic" structure is the bicontinuous gyroid phase, G, in which the minority component forms two interweaving, three-fold coordinated lattices. A perforated lamellar, or PL phase, has also been observed. It is lamellar, but with the minority-component layers perforated by "holes" that are filled by the majority component. The bicontinuous double-diamond structure has also been reported in the literature, but it is now believed that these systems were actually in G phases (Hajduk *et al.* 1995).

Nonlinear self-consistent field (NSCF) theory has been used to understand the equilibrium microphase diagrams and domain structures of neat and blended copolymers, and for a spectrum of molecular architectures (Vavasour and Whitmore 1992, 1993, Matsen and Schick 1994a, Whitmore and Vavasour 1995, Matsen and Bates 1997). In this approach, each amorphous polymer or polymer block is modeled as a Gaussian chain, with degree of polymerization Z_κ, statistical segment length b_κ, and pure component density $\rho_{o\kappa}$ which is in units of monomers per unit volume. Interactions are generally described by Flory χ parameters: finite range interactions have been incorporated in some work, but they have very little effect on the microphase diagram (Vavasour and Whitmore 1992). It is assumed that the volume of the system is independent of the degree of mixing, which is equivalent to assuming the system is incompressible.

The theory produces a set of self-consistent equations for the equilibrium density distribution of each component, and the effective field experienced by a unit at each point. For a given system, these equations are solved for each possible structure and lattice parameter (domain size), and the free energy is calculated from each self-consistent solution. By examining different structures, and a range of lattice parameters for each one, the equilibrium structure and domain size are identified.

Although experimentally all copolymer phase diagrams are qualitatively similar, they vary quantitatively from copolymer to copolymer, and they are generally not perfectly symmetric about $f_A = 0.5$. The asymmetries, and the small differences in phase diagrams for different systems for molecules of similar architectures, e.g., diblock PS-*b*-PI vs. PS-*b*-PB (polybutadiene), are largely understood in terms of a characteristic known as conformational asymmetry (Vavasour and Whitmore 1992, 1993). It represents the degree of mismatch between the volume fractions of each block and their corresponding unperturbed radii of gyration, $R_{\kappa,g}$. The asymmetry parameter can be defined as

$$\varepsilon = \left(\frac{f_A}{f_B}\right) / \left[\frac{R_{A,g}^2}{R_{B,g}^2}\right] \tag{3}$$

For conformationally symmetric polymers, $\varepsilon = 1$ and the phase diagram should be symmetric about $f_A = 0.5$. The range of ε is from about one-third to three. In defining ε, either block can be identified as the A or B block. If the choice is made so that $\varepsilon < 1$, then each microphase boundary is shifted towards higher values of f_A. This choice corresponds to the A block being the stiffer one.

A general result of the NSCF theory is that, as long as each phase of interest can be described by a single independent lattice parameter, then the entire equilibrium microphase diagram of a system of diblock copolymers depends on only three parameters, which are χZ_{eff}, f_A, and ε. Z_{eff} is an effective degree of polymerization, defined by

$$Z_{\mathrm{eff}} = \rho_0 \left[\frac{Z_A}{\rho_{0A}} + \frac{Z_B}{\rho_{0B}}\right] \tag{4}$$

where ρ_0 is the reference density used in defining the Flory χ parameter. If the two reference densities are chosen to be equal, then $Z_{\mathrm{eff}} = Z_c$.

An example of a microphase diagram calculated from NSCF theory is shown in Fig. 3 (Matsen and Bates 1997). It is for a conformationally symmetric, diblock copolymer, and so is symmetric about $f_A = 0.5$. In strong segregation, the order–order phase boundaries depend primarily on the volume fraction of each component, i.e., f_A or f_B. In the limit of very strong segregation, it is believed that only the three classical phases are stable, and the S/C and C/L phase boundaries are at limiting values of $f_A = 0.105$ and 0.310, respectively (and corresponding values at $1-f_A$) (Matsen and Whitmore 1996). At intermediate segregation, as shown in Fig. 3, the G phase becomes stable between the L and C phases, and the c.p.-S phase appears in narrow regions between the b.c.c.-S phase and the MST. In this regime, the G phase exists over a composition range of a few percent on each side of the diagram. No other ordered phase becomes stable. With decreasing χZ_c, all phase boundaries curve towards $f_A = 0.5$. The G phase terminates at triple points at $\chi Z_c = 11.14$ and $f_A = 0.452$ and 0.548 where it coexists with the L and C phases, but the L, C, and S phases all extend to the critical point where the D/S, S/C, and C/L boundaries all merge, as predicted by Leibler.

The effects of conformational asymmetry have also been examined with NSCF theory. Expressed as a phase diagram labeled by χZ_{eff} and f_A, the primary effect is shifting the microphase boundaries, with very little change to the MST itself, and no change in the topology of the phase diagram. The first calculation (Vavasour and Whitmore 1993) treated the classical phases and the case of $\varepsilon = 0.6$. In strong segregation at $\chi Z_{\mathrm{eff}} = 80$, the S_A/C_A, C_A/L, and L/C_B boundaries were all shifted by 0.07, and the C_B/S_B by 0.05. As a consequence, the width of the S_A region was approximately doubled while that of the S_B was approximately halved, whereas the others were shifted but with little changed in size. Matsen, Schick, and Bates have investigated other values of ε and the other phases (Matsen and Schick 1994b, 1994c, Matsen and Bates 1997).

They obtained similar results, with shifts in the C/G and G/L boundaries similar to those of the other microphase boundaries. The L, C, and S phases extend to the critical point, and the only direct $D \leftrightarrow L$ transition occurs at this point. However, the critical point is shifted away from $f_A = 0.5$ in the same direction as the phase boundaries and upwards to $\chi Z_{\mathrm{eff}} > 10.5$, and it no longer occurs at the minimum in the MST curve. The PL phase is nearly stable in the calculations, and Matsen and Bates suggest that it might become stable for high conformational asymmetry and $\chi Z_c \gtrsim 35$.

Matsen and Schick (1994b, 1994c) have also published phase diagrams for conformationally symmetric and asymmetric multiblock copolymers in the limit of very many blocks, conformationally symmetric starblocks with up to nine arms, and conformationally asymmetric starblocks with five arms. All these phase diagrams were topologically the same.

The current experimental picture of the microphase behavior of amorphous block copolymers has been summarized in a number of references (Bates 1991, Matsen and Bates 1997, Bates and Fredrickson 1999). As predicted by the NSCF theory, the phase boundaries are vertical in strong segregation, but they appear to remain approximately vertical right to the MST so that the phase boundaries do not merge and there are transitions directly from the disordered phase to each of the S, C, G, and L phases. The G phase appears between the C and L phases, over a composition range of a few percent, as predicted by the NSCF theory. However, it has not been observed beyond $\chi Z_c \approx 30$; instead, the PL phase is observed there. However, this phase is now thought to be metastable; whether it is stable or metastable, these observations are consistent with the NSCF result that the G and PL phases have nearly equal free energies in this region. A general conclusion is that there is very good overall NSCF/experimental agreement, except very near the MST, where the theory predicts only D/S transitions.

Since the NSCF theory can be applied from the MST to very strongly segregated systems, it is illustrative to compare it with other mean field theories which apply to either the strong or weak segregation limits. This comparison is made in Fig. 4. The right-hand side of this figure compares the NSCF results with the strong segregation theory of Helfand and Wasserman (1982), which is a numerical SCF theory but with the explicit assumption that the A–B interphase is narrow, the "narrow interphase approximation." For $\chi Z_c \geqslant 80$, the L/C and S/C boundaries agree almost perfectly, but there is still some difference in the predicted MST even here. At intermediate segregation, $20 \lesssim \chi Z_c \lesssim 45$, there is a qualitative difference: the MST is predicted to be $D \leftrightarrow C$ or $D \leftrightarrow S$ in the Helfand work. In very strong segregation, the NSCF results also agree well with the theory of Semenov (1985), although there may remain a very small unexplained discrepancy (Matsen and Whitmore 1996).

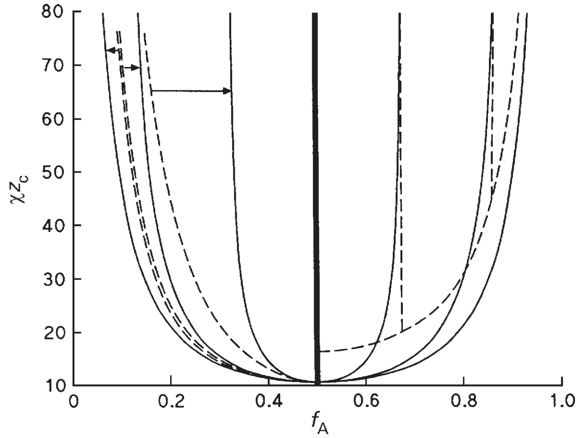


Figure 4
Comparison of the NSCF phase diagram (solid lines) with limiting theories, for conformationally symmetric, diblock copolymers (Vavasour and Whitmore 1992, 1993). For $f_A < 0.5$, the dashed curves are calculated using the weak segregation theory of Leibler (1980), with the arrows showing the corresponding phase boundaries in two approaches. For $f_A > 0.5$, the dashed curves are calculated using the narrow interphase approximation of Helfand and Wasserman (1982).

Weak segregation theories have often been used to interpret microphase separated systems near the MST, even when this is at large χZ_c. The left-hand side of Fig. 4 compares the NSCF results with Leibler's weak segregation theory (Leibler 1980). There is complete agreement as $f_A \rightarrow 0.5$ and $\chi Z_c \rightarrow 10.5$, but the domain of agreement is restricted to about $0.45 \lesssim f_A \lesssim 0.55$, and $\chi Z_c \lesssim 15$. For $\chi Z_c \gtrsim 15$, the NSCF microphase boundaries turn upwards more quickly, resulting in much broader C and S regions and a narrower L region. Even near the MST, there are important differences for asymmetric polymers. The S regions are much larger in the NSCF results, and the MST is shifted. For example, at $f_A = 0.1$ (or 0.9), NSCF theory predicts the MST at $\chi Z_c \approx 50$, whereas the Leibler theory predicts it at $\chi Z_c \approx 80$.

All mean field theories of weak segregation predict the MST to be from the disordered phase to spheres, except at the critical point. This disagreement with experimental results is thought to be due to the effects of fluctuations (Fredrickson and Helfand 1987). However, quantitative agreement between theory and experiment has yet to be achieved in this part of the phase diagram.

2.2 *Domain Sizes and Density Distributions*

In weak segregation, each density profile can be expressed as a sinusoidal variation about the mean density, containing wave vectors of only one magnitude. In the other extreme of strong segregation, each species is present in only one subdomain, and the density profiles are flat except for a smooth but narrow interphase region. Between these two limits, the profiles evolve from one shape to the other.

The NSCF theory provides a detailed account of the density profiles without making *a priori* assumptions about their shape. The results show that, for $\chi Z_c \lesssim 15$, each density profile is approximately a sine-like variation about its mean, with the amplitude of the variation increasing with increasing χZ_c. Beyond this point, the maxima and minima in each profile reach their limiting values, and the profiles begin to flatten in each subdomain. As χZ_c is progressively increased, the interphase sharpens, the profiles become more step-like, and the plateau-like regions in each subdomain extend over larger distances.

Leibler predicted that the domain size at the MST is approximately equal to the unperturbed radius of gyration, Eqn. (2). Once a microphase forms, the molecules stretch and d increases. The NSCF theory predicts that if the dependence is expressed as a power law then d scales as

$$d \propto (\chi Z_c)^p Z_c^{1/2} \tag{5}$$

throughout the phase diagram (Vavasour and Whitmore 1992). The variation of d with both χ and Z_c is actually strongest in weak segregation near the mean field critical point. This form agrees with the very strong segregation theory of Semenov (1985), which predicts $d \propto \chi^{1/6} Z_c^{2/3}$. As elaborated elsewhere (Whitmore and Vavasour 1995), a variety of NSCF calculations on diblocks, triblocks, multiblocks, and blends are all consistent with this result, with numerically calculated values of p ranging from about 1/6 in strong segregation to 0.5 at the critical point. It should be noted that Eqn. (5) applies within each structure; small discontinuities can occur at each order–order phase boundary (Matsen and Bates 1997).

The variations in the scaling of the domain size with χ and Z_c can be correlated with the density profiles. In the part of the phase diagram where the density profiles are step-like, the scaling is relatively weak, with the value of p of the order of 1/6. In regions where the profiles are sine-wave-like with smaller amplitudes, the dependence strengthens. As the MST is approached, it appears to reach a limiting effective value of $P = 1/2$ at the mean field critical point for symmetric copolymers. It does not reach this high for asymmetric copolymers because of the intervening, first order transition to the disordered phase.

These considerations also offer insight into the nature of the MST and the applicability of weak segregation theories. For $f_A \lesssim 0.2$ or $f_A \gtrsim 0.8$, microphases exist only for $\chi Z_c \gtrsim 15$. For these highly asymmetric copolymers, the density profiles are non-sinusoidal and have large amplitudes even at the MST. This contrasts with the assumptions of the weak segregation theories, and explains why the Leibler and NSCF microphase diagrams agree only for relatively symmetric copolymers near the MST.

As is often the case in polymer systems, much of the interesting behavior can be understood on the basis of a simple physical picture. In the present case, we consider a conformationally symmetric system with equal monomer reference volumes, ρ_0, and statistical segment lengths, b. In strong segregation, there are three main contributions to the free energy of a microphase separated system state relative to the homogeneous reference state,

$$\Delta f \approx \Delta f_{\text{int}} + f_I + f_{\text{el}} \tag{6}$$

where this is the free energy per unit volume in units of $k_B T \rho_0$, where k_B is the Boltzmann constant and T is temperature (in kelvin). The first contribution is the change in interaction energy resulting from the reduction in A–B contacts. In the limit of strong segregation,

$$\Delta f_{\text{int}} \approx -\chi f_A f_B \tag{7}$$

This term is the driving force for the MST to occur, but it is independent of the resulting domain size. The second is the energy associated with the interfacial tension at each A–B interphase. It can be approximated as

$$f_I \approx \gamma \frac{2}{d} \tag{8}$$

where the interfacial tension can be approximated in these units by (Helfand and Tagami 1971):

$$\gamma \approx \left(\frac{\chi}{6}\right)^{1/2} b \tag{9}$$

This contribution decreases with increasing d, and so favors large domain sizes. However, large domain sizes cause stretching of the polymer blocks. The third term, f_{el}, is the elastic free energy owing to this stretching. It can be approximated as

$$f_{\text{el}} \approx \frac{1}{2Z_c}\left[\alpha^2 + \frac{2}{\alpha} - 3\right] \tag{10}$$

where

$$\alpha = \frac{d}{Z_c^{1/2} b} \tag{11}$$

is a measure of the stretching of the polymer. Using Eqns. (7) to (11) in Eqn. (6), and minimizing Eqn. (6) with respect to d, yields

$$d \propto \chi^{1/6} Z_c^{2/3} b \tag{12}$$

$$\Delta f \approx \chi\left[-f_A f_B + \left(\frac{2}{3\chi Z_c}\right)^{2/3}\right] \tag{13}$$

Equation (13) implies that the MST should be governed by two factors, χZ_c and $f_A f_B$, and should occur at approximately

$$\chi Z_c \approx \frac{3}{2}\left(\frac{1}{f_A f_B}\right)^{3/2} \tag{14}$$

which implies that, for a given χZ_c, the MST should occur at

$$f_A = 0.5 \pm \frac{1}{2}\left[1 - \left(\frac{12}{\chi Z_c}\right)^{2/3}\right]^{1/2} \tag{15}$$

These simple equations are very accurate in strong segregation. Equation (12) agrees with the numerical results, with $P = 1/6$. For all $\chi Z_c \gtrsim 20$, Eqn. (15) predicts the MST at values of f_A that agree with the full NSCF result to within 3%. Even in weak segregation the MST is qualitatively correct: it is predicted to occur at $\chi Z_c = 12$, instead of 10.5, for $f_A = 0.5$. This accuracy supports the picture of a balance of the three physical factors identified above as controlling the transition and domain sizes, except for $\chi Z_c \lesssim$ 15. However, the free energy differences between different microphases are too small for this simple picture to discriminate between them.

3. Amorphous Copolymer/Solvent Blends

Copolymers are often dissolved in solvent or blended with one or more types of homopolymer. Broadly speaking, there are three general problems related to the phase behavior of these blends: the overall microphase behavior; the interplay between microphase and macrophase separation; and micelle formation at low copolymer concentration. Copolymer can also be used as surfactant in homopolymer/homopolymer blends, but that is not discussed here.

3.1 Nonselective Solvent

It is useful to consider first the ideal case in which the solvent/polymer Flory parameters are exactly equal for each block, $\chi_{SA} = \chi_{SB}$. An important concept for this system is the dilution approximation, which assumes that the only role of the solvent is to screen the A–B interactions, and that the solvent density remains uniform throughout the system. With this approximation, the microphase diagram and domain size behavior for concentrated solutions are the same as for neat copolymer, except that the A–B Flory parameter is replaced by an effective parameter,

$$\chi_{\text{eff}} = \phi_C \chi_{AB} \tag{16}$$

where ϕ_C is the overall copolymer volume fraction. This implies that the added solvent shifts the MST to higher $\chi_{AB} Z_c$; for example, to

$$\chi_{AB} Z_c = \frac{10.5}{\phi_C} \tag{17}$$

for $f_A = 0.5$. Furthermore, it implies that the domain size scales as

$$d \propto (\chi_{\text{eff}} Z_C)^p Z_C^{1/2} \propto (\phi_C \chi_{AB})^p Z_C^{p+1/2} \tag{18}$$

so that d varies with polymer concentration in the same way that it does with χ_{AB}, and that the dependence is strongest in weak segregation. Notice that this implies that the addition of nonselective solvent causes the equilibrium domain size to *decrease*.

The full NSCF calculations give nearly uniform solvent density profiles, as assumed in the dilution approximation. The calculations indicate only a very small solvent excess in the A–B interphase region, even in strong segregation where the polymer densities vary sharply through the interphase (Whitmore and Vavasour 1995). When expressed in terms of χ_{eff}, the microphase diagrams are essentially identical with that for neat copolymer. The layer thickness scaling was found to be qualitatively consistent with Eqn. (18) although there were small quantitative differences.

In the semidilute regime, chain swelling must be included. For conformationally symmetric polymers with $f_A = 0.5$ in nonselective, good solvents in this regime, the MST is predicted to obey (Fredrickson and Liebler 1989, Olvera de la Cruz 1989):

$$\chi_{AB} Z_c \approx \frac{10.5}{\phi_C^{1.61}} \tag{19}$$

and the layer thickness at the MST is predicted to scale as

$$d \propto Z_C^{1/2} \phi_C^{-0.115} \tag{20}$$

In the limit of very strong segregation, the order–order microphase boundaries are predicted to be unchanged in these systems (Birshtein and Zhulina 1990).

Turning to experiment, Hashimoto *et al.* (1983) and Lodge *et al.* (1995) studied PS-*b*-PI in dioctyl phthalate (DOP) and toluene, which are neutral, good solvents for PS and PI. For systems in equilibrium, the layer thickness (Hashimoto *et al.* 1983) scaled as

$$d \propto \phi_C^{1/3} \left(\frac{1}{T}\right)^{1/3} Z^{2/3} \tag{21}$$

These exponents were determined without distinguishing between strong and weak segregation, and the concentration dependence approximately followed a 1/3 power all the way to the MST; in fact, the results are consistent with a slight strengthening of the power in the neighborhood of the MST. These results are consistent with the NSCF predictions for the concentrated regime, assuming that the interaction parameters are independent of composition and inversely proportional to temperature. However, the copolymer volume fraction at the MST scaled as $\varphi_C \propto Z_c^{-0.62}$, from $\varphi_C = 0.14$ all the way up to the melt, i.e., it obeyed the scaling predicted for the semidilute regime, Eqn. (19).

3.2 Selective Solvent

The behavior of systems with slightly selective solvents is similar to that of the idealized, perfectly nonselective solvent cases just described (Banaszak and Whitmore 1992, Lodge *et al.* 1997). The solvent is partially partitioned between subdomains, but it is approximately uniform within each subdomain, and the variation between layers is relatively small. Nonetheless, even a slight difference can be sufficient to eliminate the small local maximum in each interphase, so that the solvent density decreases monotonically from one subdomain to the other. The scaling of the layer thickness with polymer concentration was calculated to be weaker for the selective solvents than for the nonselective case, but its dependence on χ_{AB} and Z_c was little changed (Banaszak and Whitmore 1992).

Huang and Lodge (1998) carried out an extensive NSCF study of the effects of solvent selectivity on the microphase diagram, for the classical structures. In general, the MST is lowered to smaller values of $\chi_{AB}Z_c$ as the solvent selectivity is increased. Put another way, the selective solvent has a weaker tendency to destabilize the microphase than does the nonselective one, and the difference can be very significant (Banaszak and Whitmore 1992). For a strongly selective solvent, the predicted order of phases with increasing polymer concentration can become complex, and the predicted sequence of phases can be $S_A \rightarrow C_A \rightarrow L \rightarrow C_B \rightarrow S_B$.

4. *Amorphous Copolymer/Homopolymer Blends*

Other interesting effects can occur in binary, ternary, or multicomponent copolymer/homopolymer blends. The copolymer itself may or may not microphase separate, and the blends can macrophase separate. Each phase can be microphase separated, or disordered. If the homopolymer is solubilized within a microdomain, it can modify the domain sizes. A full understanding of these systems requires a construction of the global phase diagram (Zin and Roe 1984, Roe and Zin 1984) and the characterization of each phase; we limit the discussion here to the MST of copolymer/homopolymer blends, and some understanding of the weak segregation regime.

The addition of low molecular weight homopolymer has similar effects as solvent: it tends to destabilize the microphase and reduce the domain sizes. However, high molecular weight homopolymer has opposite effects: when added to disordered copolymer near the MST it can cause the system to microphase separate with the homopolymer solubilized within the domain structure. This is called induced microphase formation (Noolandi and Hong 1983, Cohen and Torradas 1984).

Not only do low and high molecular weight homopolymer have opposite effects on the stability of the microphase, they can also have opposite effects on the domain size. Experimentally (Quan *et al.* 1987, Hashimoto *et al.* 1990, Tanaka and Hashimoto 1991, Tanaka *et al.* 1991, Winey *et al.* 1991), the solubilization of very low molecular weight homopolymer causes either a decrease or very slight increase in the domain size. The addition of high molecular weight homopolymer causes an increase in domain size, up to the solubilization limit of the homopolymer.

A detailed analysis based on weak segregation theory provides a qualitative, physical picture of the behavior of these blends. It predicts two threshold effects associated with homopolymer added to neat copolymer systems. For A-*b*-B/A binary blends with symmetric copolymers, the threshold for homopolymer-induced microphase formation should occur near

$$\frac{Z_h}{Z_c} \approx \frac{1}{4} \tag{22}$$

where Z_h and Z_c are the degrees of polymerization of the homopolymer and copolymer, respectively. Added homopolymer with $Z_h \lesssim Z_c/4$ tends to destabilize a microphase, whereas if $Z_h \gtrsim Z_c/4$, it tends to stabilize, or induce, the microphase. The threshold effect associated with the layer thickness for these blends occurs at

$$\frac{Z_h}{Z_c} \approx \frac{1}{5} \tag{23a}$$

The work on the effects of solubilized homopolymer on weakly segregated copolymer has now been generalized to all segregation regimes (Vavasour and Whitmore 2001). The threshold associated with an increase or decrease in layer thickness, Eqn. (23a), becomes

$$\frac{Z_h}{Z_c} \approx \frac{1}{\chi Z_c}(1.39\phi_c + 0.68) \tag{23b}$$

where ϕ_c is the copolymer volume fraction to which the homopolymer is added. As implied by Eqn. (23b), the threshold decreases with increasingly strong copolymer segregation. This result is in good agreement with experiments, and reduces to Eqn. (23a) when $\phi_c = 1$ and $\chi Z_c = 10.5$.

Added homopolymer with $Z_h \lesssim Z_c/5$ causes the layer thickness to decrease, whereas it increases if $Z_h \gtrsim Z_c/5$. These results are in qualitative agreement with the experiments described above. For *ternary* A-*b*-B/A/B blends, this theory predicts homopolymer-induced microphase separation, with the same threshold as for binary blends, Eqn. (22), independent of the relative volume fractions of A and B homopolymer.

The similarity of these two thresholds, and the calculated density profiles, provides an overall picture of the effects of these additives. Adding solvent or low molecular weight homopolymer tends to destabilize the microphase, reduce the domain thickness, and move the system towards weaker segregation with density profiles becoming more sinusoidal with smaller amplitudes. High molecular weight homopolymer tends to stabilize the microphase, cause *d* to increase, and drive the system towards stronger segregation and more step-like profiles. This is ultimately limited by the solubilization limit of the homopolymer.

5. *Block Copolymer Micelles*

Another form of microphase separation of block copolymers is the formation of micelles when low concentration copolymer (a few percent or less) is added to a selective solvent or homopolymer, such that one block of the copolymer is compatible with the host, but the other is not. The micelles are usually approximately spherical, and of the order of 10–30 nm in size.

Each core consists primarily of the incompatible copolymer blocks, while the compatible blocks form a corona extending into the host.

Interesting questions in these systems include the conditions under which micelles form, i.e., the critical micelle concentration (cmc), and the size and number density of the micelles. Although nonequilibrium effects can, again, be very important, simple equilibrium theories provide insight into the physics governing these systems, and good descriptions of systems which are thought to be in equilibrium (Leibler *et al.* 1983, Whitmore and Noolandi 1985). These models are similar in essence to the physical model introduced at the end of Sect. 2.

As in Eqn. (6), we consider the free energy of a system of micelles relative to a reference state of copolymer dissolved in the host. We can express the important contributions as

$$\Delta f \approx \Delta f_{\text{int}} + f_{\text{I}} + f_{\text{el}} + f_{\text{loc}} \tag{24}$$

The first three terms are similar to those in Eqn. (6). The first, Δf_{int}, is the reduction in interaction energy when the incompatible copolymer blocks avoid contact with the host by forming the micelle cores. The second term, f_{I}, is the interfacial energy, in this case arising from the core-corona interface. The third, f_{el}, is the stretching term, which differs for the core and corona blocks but has the same functional form for each one. The final term, f_{loc}, is due to the reduction in entropy that occurs when the copolymers are localized to micelles.

As in the copolymer model of the MST, Δf_{int} is the driving force for forming micelles, but it does not control the micelle sizes. These are, again, controlled primarily by the balance between the interfacial tension and stretching, but this time the stretching of the core block. As a result, for a simplified model system consisting of A-*b*-B copolymer in an A-homopolymer or A-block compatible solvent, such that a single χ parameter describes all unfavorable interactions, the equilibrium core radius is predicted to scale approximately as

$$l_{\text{core}} \approx \chi^{1/6} Z_{\text{cB}}^{2/3} b \tag{25}$$

This is similar scaling behavior as for microphase separated copolymers. The main difference is that the micelle core radius scales with the degree of polymerization of the corresponding copolymer *block*, Z_{cB}. The cmc behaves approximately as

$$\phi_{\text{C}}^{\text{cmc}} \propto \frac{\chi_{\text{AB}} Z_{\text{c}}}{(\chi_{\text{AB}} Z_{\text{c}})^{2/3}} \mathrm{e}^{\{-\chi_{\text{AB}} Z_{\text{c}} + \cdots\}} \tag{26}$$

This expression indicates that the onset of micellization is dominated by the exponential dependence on χZ_{cB}, a rather different condition to Eqn. (1) for copolymer phase separation. It should be noted that although Eqns. (25) and (26) capture the dominant physical effects; more quantitative results are available in the literature (Leibler *et al.* 1983, Whitmore and Noolandi 1985).

6. *Semicrystallizable Diblock Copolymers*

Semicrystallizable diblock copolymers have one crystallizable and one amorphous block. These systems tend to form equilibrium lamellar structures, in which one sublayer consists of folded, semicrystalline blocks, and the other one is amorphous. The structure is illustrated schematically in the lower panel of Fig. 1. An example is PS-*b*-PEO, in which the PEO block is the semicrystalline one.

As with amorphous copolymers, the equilibrium layer thickness is governed by a balance of thermodynamic driving forces. However, the underlying physics differs. It, too, can be understood on the basis of an NSCF theory (Whitmore and Vavasour 1995). In this theory, each layer of thickness d contains amorphous and semicrystalline subdomains of thickness d_{A} and d_{B}, respectively. All the A–B joints are assumed to be localized within a narrow interphase of thickness a and, except for this interphase, all A and B monomers are assumed to be in their respective subdomains. The amorphous A blocks are treated in the usual way. However, for the B blocks, the energetics associated with crystallization and chain folding are explicitly included instead. These blocks are characterized by the heat of fusion per unit volume, ΔH_{f}, the degree of crystallinity of the B block, τ_{c}, and the energy of each fold, E_{fold}.

The free energy density, relative to a hypothetical uniform melt, can be expressed as

$$\Delta f = f_{\text{int}} + f_{\text{loc}} + f_{\text{cr}} + f_{\text{am}} \tag{27}$$

The first term, f_{int}, is the difference in the interaction energy, and is taken to be the same as for amorphous copolymers. The next term, f_{loc}, is due to the change in entropy associated with the localization of the joints. The third term, f_{cr}, is the crystallization energy of the B block. Invoking the chain folding model, it can be expressed

$$f_{\text{cr}} = \frac{f_{\text{B}}}{k_{\text{B}} T}\left[-\tau_{\text{c}} \frac{\Delta H_{\text{f}}}{\rho_{0\text{B}}} + \frac{n_{\text{f}}}{Z_{\text{cB}}} E_{\text{fold}}\right] \tag{28}$$

where n_{f} is the number of folds per molecule. This expression, and others in this section, assume that the reference densities of both components are equal. The general case is considered elsewhere (Whitmore and Vavasour 1995).

The final term in Eqn. (27), f_{am}, is that part of Δf which can be associated solely with the amorphous region, and it is calculated via an NSCF calculation.

The result can be expressed as

$$f_{\mathrm{am}}=\frac{1}{Z_{\mathrm{c}}}g_{\gamma}(\alpha) \tag{29}$$

where $g_\gamma(\alpha)$ is the free energy associated with the amorphous region *per molecule*. It is a function of only two variables:

$$\alpha=\left(\frac{3}{Z_{\mathrm{cA}}}\right)^{1/2}\frac{d_{\mathrm{A}}}{b_{\mathrm{A}}} \tag{30}$$

which is a measure of the thickness of the amorphous region relative to unstretched A blocks, and

$$\gamma=\left(\frac{1}{Z_{\mathrm{cA}}}\right)^{1/2}\frac{a}{b_{\mathrm{A}}} \tag{31}$$

which is a measure of the thickness of the interphase region relative to unstretched A blocks.

This function g can be used for the amorphous block of any semicrystallizable copolymer. For a given value of γ, it is large both for small α, which corresponds to compressed chains, and for large α, which corresponds to stretched chains. In between, there is a minimum that occurs when the root mean squared thickness of the amorphous block is comparable with its unperturbed end-to-end distance. For a given degree of stretching, g increases with decreasing γ, which can be understood as a result of the conformational constraints imposed on the molecules by a narrower interphase. The calculated values of g can be fitted to

$$g\approx\frac{A}{\sqrt{\gamma}}\alpha^{\mu}+\frac{B}{\alpha\gamma}+C \tag{32}$$

where the constant C is small. The calculated values of μ vary from $\mu\approx2.34$ for $\gamma=0.2$ to $\mu\approx2.11$ for $\mu=0.05$. A simple extrapolation to $\gamma\rightarrow0$, which is equivalent to $Z_{\mathrm{cA}}\rightarrow\infty$ for a fixed interphase width, gives $\mu\approx2.0$, in agreement with limiting theories (Semenov 1985).

The equilibrium layer thickness, d, is obtained by minimizing the free energy (Eqn. (27)). All four of the contributions to Δf depend on d, but the dependence of two of them, f_{loc} and f_{int}, is very weak. Accordingly, the equilibrium thickness is determined primarily by a balance between f_{cr} and f_{am}.

The variation of f_{cr} with d can occur through two mechanisms. The first would be any systematic behavior of the degree of crystallinity. This would be straightforward to include, but experiments have not indicated any such systematics, so we neglect it here. The second mechanism is the chain folding. Folding occurs at the amorphous/crystalline interphases, and can occur in the interiors of the crystalline sublayers. However, these interior folds result in higher free energy, and so should not occur in equilibrium. In either case, the number of times the crystalline block traverses the subdomain scales as

$$n\propto\frac{Z_{\mathrm{cB}}}{d_{\mathrm{B}}} \tag{33}$$

The number of folds per molecule is $n_{\mathrm{f}}=n-1$.

Returning to Eqns. (28) and (29) for f_{cr} and f_{am} and assuming τ_{c} is constant, the dominant contributions to the free energy density that depend on d can be written

$$f(d)\approx\frac{1}{Z_{\mathrm{c}}}\left[n_{\mathrm{f}}\left(\frac{E_{\mathrm{fold}}}{k_{\mathrm{B}}T}\right)+g_{\gamma}(\alpha)\right] \tag{34}$$

This expression illustrates the essential physics of the problem. The number of folds, and hence the free energy of the semicrystalline block, are reduced if the layer thickness increases, but this tendency is opposed by the associated reduction in entropy of the stretched amorphous blocks.

Since the amorphous blocks are stretched at equilibrium, it is reasonable to approximate g by the leading term, $g\propto\alpha^{\mu}/\sqrt{\gamma}$. Using this in Eqn. (34) and minimizing, yields a scaling relation for the equilibrium domain thickness. In the limit of $\mu\rightarrow2$, it becomes

$$d_{\mathrm{eq}}\propto\left(\frac{E_{\mathrm{fold}}}{k_{\mathrm{B}}T}\right)^{1/3}\frac{Z_{\mathrm{c}}}{Z_{\mathrm{cA}}^{5/12}} \tag{35}$$

This result is quite different from the corresponding prediction for amorphous polymers. Also, the equilibrium number of folds scales as

$$n_{\mathrm{f,eq}}\propto\left(\frac{k_{\mathrm{B}}T}{E_{\mathrm{fold}}}\right)^{1/3}Z_{\mathrm{cA}}^{5/12} \tag{36}$$

This implies that the equilibrium number of folds of the *crystalline* block depends on the degree of polymerization of the *amorphous* block, Z_{cA}, and is independent of Z_{cB}.

Birshtein and Zhulina (1990) developed an analytic theory of semicrystallizable copolymers for large Z_{cA}, based on the Semenov approach. They also predicted that the number of folds depends only on Z_{cA} but, because their free energy for the amorphous block is $g\propto\alpha^2$, independent of γ, they obtained a value for the power of 1/3 instead of 5/12.

Turning to experiment, Douzinas *et al.* (1991) studied eight (ethylene-co-butylene)-ethylethylene copolymers with total and amorphous block degrees of polymerization ranging from about 760 to 2900, and 110 to 1600, respectively. They found that the equilibrium thickness scaled as $Z_{\mathrm{c}}/Z_{\mathrm{cA}}^{\nu}$, with a best fit value for the power of $\nu=0.42\pm0.02$, agreeing with the theoretical value of $5/12\approx0.417$ to within experimental accuracy. Rangarajan and Register performed a similar study of seven ethylene-(ethylene-*alt*-propylene) copolymers (Rangarajan *et al.* 1993).

Their best fit gave a slightly stronger inverse dependence, $d \propto Z_c/Z_{cA}^{0.45}$. They speculated that the small difference might be owing to the presence of ethyl branches in the E block.

7. Concluding Remarks

Our understanding of the equilibrium microphase behavior of block copolymers, and certain aspects of block copolymer blends, is well advanced. The formation of a microphase can be understood on the basis of simple physical pictures and thermodynamic arguments. In the case of amorphous polymers, and amorphous copolymer micelles, the effective interaction between unlike monomers is the driving force for the transition. The transition is opposed by the resulting reduction in entropy. In the case of semicrystallizable copolymers, the driving force is the heat of crystallization of the corresponding block.

The domain sizes can also be understood by simple arguments. In the case of strongly segregated amorphous polymers, they are controlled by the balance of interfacial energy at the A–B interphases, and the stretching free energy of the blocks. In the micelle case, it is the stretching of the core-forming block which is important in this balance. In the semicrystallizable copolymers, it is predominantly a balance between the stretching of the amorphous block and the energy of the folds in the crystallizable block.

In copolymer/homopolymer blends, these balances introduce new phenomena. Addition of low molecular weight homopolymer tends to destabilize a microphase, whereas relatively high molecular weight has the opposite effect. Theoretical predictions of the threshold for these effects are in qualitative agreement with experimental.

The free energy differences between different microphases are small, and it is difficult to understand them on the basis of simple physical pictures. Numerical self-consistent field theory does well, except near the MST. It is likely that the incorporation of fluctuation effects into a full numerical SCF theory will be necessary for a full understanding of this complex region of the phase diagram.

Bibliography

Banaszak M, Whitmore M D 1992 Self-consistent theory of block copolymer blends: selective solvent. *Macromolecules* **25**, 3406–12

Bates F S 1991 Polymer–polymer phase behavior. *Science* **241**, 845

Bates F S, Fredrickson G H 1999 Block copolymers—designer soft materials. *Phys. Today* **52** (2), 32–8

Birshtein T M, Zhulina E B 1990 Scaling theory of super molecular structures in block copolymer–solvent systems. 2. Super crystalline structures. *Polymer* **31**, 1312–20

Cohen R E, Torradas J M 1984 Homopolymer-induced microphase separation of a homogeneous diblock copolymer. *Macromolecules* **17**, 1101–2

Douzinas K C, Cohen R E, Halasa A F 1991 Evaluation of domain spacing scaling laws for semicrystalline diblock copolymers. *Macromolecules* **24**, 4457–9

Fredrickson G H, Helfand E 1987 Fluctuation effects in the theory of microphase separation in block copolymers. *J. Chem. Phys.* **87**, 697–705

Fredrickson G H, Leibler L 1989 Theory of block copolymer solutions: nonselective good solvents. *Macromolecules* **22**, 1238–50

Hajduk D A, Harper P E, Gruner S M, Honeker C C, Thomas E L, Fetters L J 1995 A re-evaluation of bicontinuous cubic phases in starblock copolymers. *Macromolecules* **28**, 2570–3

Hashimoto T, Shibayama M, Kawai H 1983 Ordered structure in block copolymer solutions. 4. Scaling rules on size fluctuations with block weight, concentration, and temperature in segregation and homogeneous regimes. *Macromolecules* **16**, 1093–104

Hashimoto T, Tanaka T, Hasegawa H 1990 Ordered structure in mixtures of a block copolymer and homopolymer. 2. Effect of molecular weight of homopolymers. *Macromolecules* **23**, 4378–86

Helfand E, Tagami Y 1971 Theory of the interface between immiscible polymers. II. *J. Chem. Phys.* **56**, 3592–601

Helfand E, Wasserman Z R 1982 Microdomain structure and the interface in block copolymers. In: Goodman I (ed.) *Developments in Block Copolymers*. Elsevier, New York, Vol. 1, pp. 99–125

Huang C -I, Lodge T P 1998 Self-consistent calculations of block copolymer solution phase behavior. *Macromolecules* **31**, 3556–65

Leibler L 1980 Theory of microphase separation in block copolymers. *Macromolecules* **13**, 1602–17

Leibler L, Orland H, Wheeler J C 1983 Theory of critical micelle concentration for solutions of block copolymers. *J. Chem. Phys.* **79**, 3550–7

Lodge T P, Hamersky M W, Hanley K J, Huang C -I 1997 Solvent distribution in weakly-ordered block copolymer solutions. *Macromolecules* **30**, 6139–49

Lodge T P, Pan C, Jin X, Liu Z, Zhao J, Maurer W W, Bates F S 1995 Failure of the dilution approximation in block copolymer solutions. *J. Polymer Sci., Polymer Phys. Ed.* **33**, 2289–93

Matsen M W, Bates F S 1997 Conformationally asymmetric block copolymers. *J. Polymer Sci. B* **35**, 945–52

Matsen M W, Schick M 1994a Stable and unstable phases in a diblock copolymer melt. *Phys. Rev. Lett.* **16**, 2660–3

Matsen M W, Schick M 1994b Microphase separation in starblock copolymer melts. *Macromolecules* **27**, 6761–7

Matsen M W, Schick M 1994c Stable and unstable phases in a linear multiblock copolymer melt. *Macromolecules* **27**, 7157–63

Matsen M W, Whitmore M D 1996 Accurate diblock copolymer phase boundaries at strong segregations. *J. Chem. Phys.* **105**, 9698–701

Noolandi J, Hong K M 1983 Theory of phase equilibrium in systems containing block copolymers. *Macromolecules* **16**, 1083–93

Olvera de la Cruz M 1989 Theory of microphase separation in block copolymer solutions. *J. Chem. Phys.* **90**, 1995–2002

Quan X, Gancarz I, Koberstein J T, Wignall G D 1987 Effect of homopolymer molecular weight on the morphology of block copolymer/homopolymer blends. *Macromolecules* **20**, 1431–4

Rangarajan P, Register R A, Fetters L J 1993 Morphology of semicrystalline block copolymers of ethylene-(ethylene-alt-propylene). *Macromolecules* **26**, 4640–5

Roe R -J, Zin W-C 1984 Phase equilibrium and transition in mixtures of a homopolymer and a block copolymer. 2. Phase diagram. *Macromolecules* **17**, 189–94

Semenov A N 1985 Contribution to the theory of microphase layering in block-copolymer melts. *Sov. Phys-JETP* (English translation) **61**, 733–742

Tanaka H, Hasegawa H, Hashimoto T 1991 Ordered structure in mixtures of a block copolymer and homopolymer. 1. Solubilization of low molecular weight homopolymers. *Macromolecules* **24**, 240–51

Tanaka H, Hashimoto T 1991 Ordered structures of block polymer/homopolymer mixtures. 3. Temperature dependence. *Macromolecules* **24**, 5713–20

Vavasour J D, Whitmore M D 1992 Self-consistent mean-field theory of the microphases of diblock copolymers. *Macromolecules* **25**, 5477–86

Vavasour J D, Whitmore M D 1993 Self-consistent mean-field theory of block copolymers with confrontational asymmetry. *Macromolecules* **26**, 7070–5

Vavasour J D, Whitmore M D 2001 Effects of solubilized homopolymer on lamellar diblocks. Macromolecules (in press)

Whitmore M D, Noolandi J 1985 Theory of micelle formation in block copolymer–homopolymer blends. *Macromolecules* **18**, 657–65

Whitmore M D, Vavasour J D 1995 Self-consistent field theory of block copolymers and block copolymer blends [A Review]. *Acta Polym.* **46**, 341–60 (This reference contains an extensive bibliography of literature up to 1995)

Winey K I, Thomas E L, Fetters L J 1991 Swelling a lamellar diblock copolymer with homopolymer. Influences of homopolymer concentration and molecular weight. *Macromolecules* **24**, 6182–8

Zin W -C, Roe R-J 1984 Small-angle x-ray scattering study. *Macromolecules* **17**, 183–8

M. D. Whitmore

Memorial University of Newfoundland, St. John's Newfoundland, Canada

Block Copolymers, Structural Characterization of

In the morphological characterization of block copolymers, one is typically interested in obtaining the following types of basic information: (i) the geometry of the microphase-separated domains (i.e., spheres, cylinders, lamellae, perforated lamellae, bicontinuous structure); (ii) the lattice type and crystallographic space group if the structure has long-range order; and (iii) the dimensions of the lattice and of the domains ordered on the lattice. In addition to this basic information, one may also be interested in obtaining information on the mechanism and kinetics of microphase separation or order–order transitions, or on defects in the lattice structure.

Whenever possible, it is important to obtain both real space (microscopy) and reciprocal space (scattering) information with which to evaluate the microphase-separated block copolymer structure. The most common microscopy technique used with block copolymers is transmission electron microscopy (TEM). Small-angle x-ray scattering (SAXS) and small-angle neutron scattering (SANS) are the reciprocal space methods of choice. This article will focus on the application of these characterization techniques. TEM and scattering techniques provide complementary information. TEM provides direct visual clues to the geometry of the structure: but its sampling statistics are poor, it is not an accurate means of measuring dimensions, and it is subject to artifacts as a result of thin sectioning and the projection mechanism of image formation. In contrast, scattering techniques yield excellent statistics and accurate structural dimensions, but may be difficult to interpret without visual clues from TEM images.

1. *Transmission Electron Microscopy*

TEM has the advantage of producing real space images which may be directly interpretable, provided one keeps in mind the mechanism by which contrast arises, as well as how thin sectioning and projection of electrons through the sample generate the image. Figure 1 shows a TEM image of a poly(styrene-*b*-butadiene) block copolymer which has been stained with OsO_4, rendering the polybutadiene domains dark.

Due to the extremely strong interaction between electrons and matter, samples for TEM observation must be very thin, typically 1000 Å or less for block copolymers. The total areas of thin sample typically observed in the magnification range commonly employed (about ×5k to ×50k) are minuscule. Thus the data are representative of only a very small volume of the sample, and should be trusted only when consistent with scattering data (obtained by sampling a much larger volume). Scattering techniques can provide some information about defect structures in block copolymer materials, specifically grain size information. However, detailed studies of defect morphology require TEM imaging without the possibility of corroborating information from scattering, since the structures of interest represent only a tiny fraction of the material. In this case, special care should be used to employ a "tilt series," in which a number of images of the same area of sample are taken from different projection directions and compared to simulation results.

Bombarding an organic material with electrons is inherently very destructive, breaking bonds and producing free radicals. The effect of these processes on the morphology observed in the microscope varies based on the block materials present and on the staining procedures used. Some materials, such as

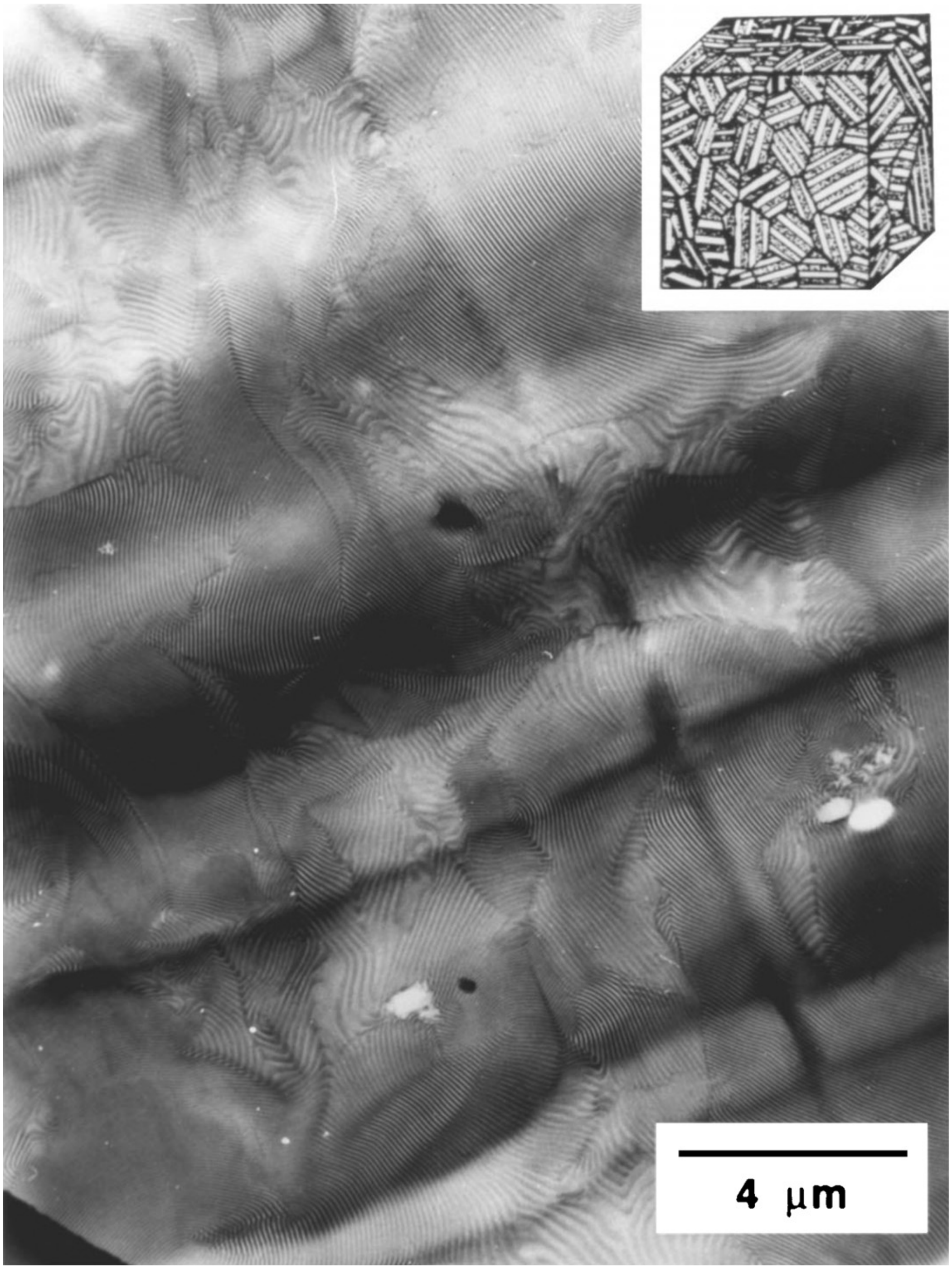


Figure 1
TEM micrograph of a multigrained lamellar structure in a poly(styrene-*b*-butadiene) diblock copolymer: PS 81000 $\mathrm{g\,mol^{-1}}$; PB 74400 $\mathrm{g\,mol^{-1}}$. Microphase-separated PB lamellae appear dark due to staining with OsO_4. The larger length scale diffuse banding in the image is the result of sample thickness variations due to microtome chatter. The inset schematically illustrates the three-dimensional grain structure of the material.

polymethylmethacrylate (PMMA), depolymerize under electron beam irradiation, producing volatile fragments. If PMMA forms the microphase-separated spheres, cylinders or bicontinuous channels, its disappearance in the TEM provides a convenient contrast mechanism, whereas if it forms the matrix phase, its degradation may result in loss of sample integrity. Polystyrene (PS), one of the most common block materials, is quite stable in the electron beam, due to the radical stabilizing effect of the phenyl rings. Diene-containing block materials (e.g., polyisoprene (PI) and polybutadiene (PB)) are usually stained with OsO_4, which reacts with the double bonds in the backbone cross-linking the chains. This staining provides mass thickness contrast that renders the diene blocks dark, relative to other block materials in TEM images. The staining reaction is known to be very efficient; essentially, every diene double bond in a thin section of block copolymer is involved in a cross-link after about four hours' exposure to OsO_4 vapors. This heavy cross-linking has the added advantage that it greatly enhances the stability of the sample in the electron beam. Thin sections of PS–PI block copolymers stained with OsO_4 are stable enough to be exposed continuously to an electron beam for hours. Such long observations of the same area may be necessary when collecting a number of different projections of the same area as in electron tomography.

Samples for TEM must be ultrathin, and thus sample preparation frequently involves ultramicrotome sectioning of block copolymers with a diamond knife, often at cryogenic temperatures. Thin samples may also be prepared by direct casting of ultrathin films from solution. The interior of the TEM column is under high vacuum, and thus samples must be free of liquid/volatile solvents. For these reasons, TEM is not well-suited to *in situ* observations of structures or to time-resolved observations of dynamics. One can only do time-resolved studies when the sample is quenched or immobilized at various stages of a process, and then observed after the fact in the TEM. It is possible to cryogenically quench samples that contain certain solvents, such as water, and then observe these vitrified samples in the TEM. This technique is common in biological applications, but has only been used in rare instances to observe block copolymer micellar structures in solution.

A central problem in the electron microscopy of polymers is obtaining contrast between different regions of the sample, i.e., between the microphase-separated domain materials in block copolymers. In typical block copolymers in which both blocks are amorphous, the only possible mechanism is mass-thickness contrast. Most electron scattering by the atoms is in the forward direction or at very small angles. As the masses of the scattering atoms increase, the probability of scattering at higher angles increases. Additionally, increasing the path of the electrons through the sample increases the probability of higher angle scattering. Use of an objective aperture to block out higher angle electron scattering from regions of the sample which are either thicker, or contain heavier elements than other regions, results in a lack of electron intensity reaching the parts of the magnified image corresponding to the highly scattering material. Thus, these regions appear darker in TEM images. In mass thickness contrast the intensity, I, at a given point in the image is given by $I = I_o \exp(-S\rho t)$, where I_o is the intensity of the incident electron beam on the sample, and ρt is the product of average density and sample thickness along a ray tracing a path through the region of the sample corresponding to the point on the image. S is a sample-specific constant of secondary importance to the relative contrast between regions of the sample. The relationship between a block copolymer morphology and its image in the TEM can be understood by ray tracing, in which the thickness and average density along many paths through a sample are evaluated and used to plot a simulated image. In doing this calculation in a block copolymer with two domain materials, it is helpful to approximate the equation for intensity by its series expansion: $I \approx I_o\ (1 - S\rho_1 t_1 - S\rho_2 t_2 + \ldots)$, where the effects on contrast of the densities and path lengths in the two different block materials are clearly delineated in the first order approximation. Sophisticated software has been developed for building models of block copolymer structures and simulating different projected TEM images by ray tracing (Anderson *et al.* 1992).

In Fig. 1 some thickness contrast occurs, due to a phenomenon known as microtome chatter, which produces long-wavelength, micron-size undulations in the thickness of microtomed sections. This results in large diffuse bands in the image. At a much smaller, nanometer-length scale, distinctive lamellar stripes are visible. These result from the higher average density of the PI domains, which contain heavy osmium metal atoms—due to staining—relative to the unstained PS. The lamellar bands visible in the image appear to have a range of thicknesses, even though the lamellar repeat distance of a near-monodisperse block copolymer material is constant. This is a projection artifact due to the orientation of the lamellar layers within the thin microtomed section. When the lamellar layers are exactly perpendicular to the thin section, the image shows their true thickness and contrast is at a maximum. As the lamellae become tilted in the film, the repeat spacing of banding in the image increases and image contrast decreases. Contrast disappears entirely when the lamellae are lying parallel to the surfaces of the thin section and thus perpendicular to the electron beam.

TEM images of other morphologies such as b.c.c. spheres, hexagonally-packed cylinders, and cubic bicontinuous gyroid structures are much more complex, but can be understood by keeping the principles of ray tracing in mind. Cylinders produce images that look like a hexagonal arrangement of circles if the projection direction is down the cylinder axis, and images which look like alternating stripes if the projection is perpendicular to the cylinder axis. This can be confusing, since the stripe projection of cylinders could be mistaken for lamellae, and the circle projection could be mistaken for spheres. Cubic packings of spheres look like a hexagonal arrangement of circles if the projection direction is down the body diagonal of the cubic unit cell. Projections parallel to the edge of the cubic sphere unit cell produce an image with a square arrangement of circles. More irregular projections of these structures are often very difficult to interpret, and care must be taken in assigning a morphology to a sample based on TEM images. In a multigrained sample, different regions of a thin section contain domains oriented in different directions, and thus a number of different projections should be visible in a single sample. This allows more accurate interpretation of the structure. Additionally, the sample may be tilted relative to the electron beam, thus changing the projection geometry. The maximum sample tilt possible is generally about $\pm 60^\circ$, but is considerably less on some TEM instruments. With a high-tilt instrument one can actually obtain, for instance, both hexagonal and square projections of the same volume of b.c.c. spherical morphology by tilting the sample 55° between the [111] direction and the [100] direction.

Staining increases contrast by depositing a heavy-element-containing species in one block domain preferentially over others. OsO_4 has already been mentioned as a stain for diene-based block materials. OsO_4 oxidizes the main chain double bonds of PI and PB, but it is not sufficiently reactive to appreciably stain the aromatic rings of PS blocks. Thus OsO_4 is an ideal stain for PS–PI and PS–PB block copolymers. RuO_4 is a stronger oxidizing agent that can also react with the aromatic unsaturation of PS. It is not a selective stain for PS–PI or PS–PB, since it stains both blocks, but it is used to selectively stain PS in block copolymers with other block materials. Poly(vinylpyridine) blocks may be selectively stained with methyl iodide. Staining can produce a modest change in domain dimensions (of the order of 5%) due to extensive chemical cross-linking. Along with the poor sampling statistics of TEM, this is another reason why scattering should be used to obtain accurate structural dimensions.

Block copolymer materials have been produced which contain significantly heavier elements in one block, such as poly(ferrocenyldimethylsilane) blocks and silicon-containing blocks. The difference between row two silicon and row one carbon, nitrogen, and oxygen is sufficient to provide useable imaging contrast in TEM. In block copolymers in which one block can crystallize within its domain structure as the temperature is lowered, the greater density and scattering power (diffraction contrast) of the crystalline material can also be used to obtain contrast between domains. This contrast will disappear in about one minute, however, as electron beam damage destroys the crystallinity. The difference in transport properties between amorphous and semicrystalline block materials has also been used to obtain contrast in polyolefin-based block copolymers by staining with RuO_4.

The incident electrons in TEM interact more strongly with the sample at lower accelerating voltages than at higher accelerating voltages. This produces greater image contrast at lower accelerating voltages. Generally, the best TEM images of block copolymer morphologies are obtained with accelerating voltages between 80 and 120 kV. Many laboratories are equipped with instruments capable of 200 kV or even greater accelerating voltages; these instruments have the advantages of greater ultimate resolution, lower beam damage, and greater ability to penetrate thicker samples. However, the advantages are of little value in block copolymer work, where the main issue is imaging contrast. The nanometer-length scale of block copolymer structure does not require high resolution, most samples are moderately to very beam stable, and thin sections can be readily prepared.

Although it is not recommended to use TEM as a means of measuring block copolymer structural dimensions, one would still like to know the scale in a TEM image as accurately as possible. In order to determine length scales in a TEM image, various size standards can be employed. The best approach is to use an internal standard such as colloidal polystyrene nanospheres. A low concentration of these commercially available particles of a very uniform, known size can be deposited on sample grids prior to imaging. The particles are available in sizes ranging from 20 nm to 10 μm. The spheres are then visible in the TEM images, and their known dimensions can be used to calibrate the magnification. Because the calibration spheres are directly in the image, they continue to provide a calibration when photographic or digital enlargements of images are produced. Another magnification calibration approach is to use commercially available TEM samples with lithographically etched grid patterns of a known size. These size standards can be put into the instrument immediately after a polymer sample and images recorded at the same magnification.

2. *Small-angle Scattering*

The volume of material that contributes to the signal in a typical small-angle x-ray (SAXS) or neutron (SANS) scattering experiment on block copolymer materials is one million or more times larger than the volume of material observed in a typical TEM image. Thus for the purposes of establishing structural symmetry or obtaining domain dimensions, the results of small-angle scattering are much more reliable than

TEM. Additionally, the staining procedures which can alter domain dimensions in TEM are not required for small-angle scattering. However, SANS contrast is greatly improved by selectively deuterating one of the block materials. The electron beam irradiation in TEM induces chemical changes such as chain scission or cross-linking; such damage generally does not occur in x-ray or neutron beams. Thus small-angle scattering techniques may be used in time-resolved or temperature-resolved experiments to observe structural changes, such as the order–disorder transition and order–order transitions *in situ* (see *Block Copolymer Phase Behavior*).

The main drawback of small-angle scattering techniques is that interpretation of the data is model-dependent; one must propose a structure and compare the expected scattering from that structure to the experimental data. This procedure is fairly straightforward for simple block copolymer morphologies (spheres, cylinders, lamellae) on well-ordered lattices, but it becomes extremely difficult when a new structure is encountered, as was the case with the cubic bicontinuous morphology. Initially, the ordered bicontinuous double diamond (OBDD) structure was proposed, but subsequently it was found that improved SAXS data could be better explained by the cubic gyroid morphology with $Ia\bar{3}d$ symmetry.

The structure of a material which is ordered on a lattice can be efficiently described by specifying a lattice symmetry which indicates how unit cells of the structure are repeated throughout space, and then by specifying how material is arranged within the unit cell. Mathematically, the structure is the convolution of the lattice with the unit cell. The resulting experimental scattering intensity is proportional to the squared amplitude of the Fourier transform of the material structure (electron density distribution). By the convolution theorem, this is equivalent to the product of the squares of the Fourier transforms, taken separately, of the lattice and the structure of the unit cell. The Fourier transform of the lattice results in the reciprocal lattice that gives the Bragg scattering peaks. The transform of the unit cell structure gives the structure factor or form factor. This contribution modulates the intensity of the observed Bragg peaks, and can also yield diffuse maxima (form factor maxima) in the observed scattering. The use of the term "form factor" with block copolymers comes from the more general field of small-angle scattering, where many problems involve trying to determine the size and shape of heterogeneous particles such as platelets or needles randomly distributed in a medium. In this case, the single-particle scattering depends on the shape or form of these particles. In block copolymers with long-range lattice order, the appropriate "form factor" contribution is the Fourier transform of the contents of the unit cell, essentially the same as the structure factor used in atomic length scale crystallography.

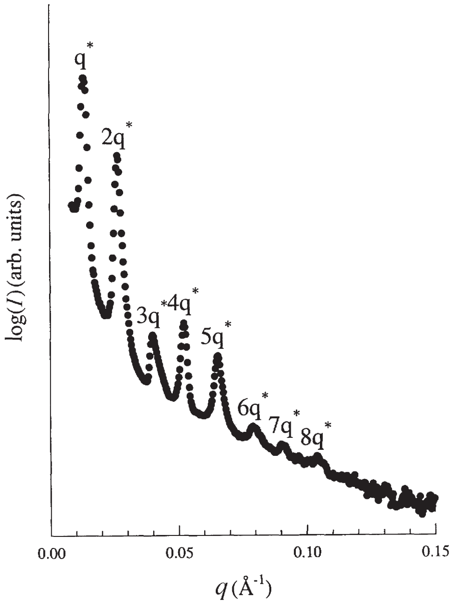


Figure 2
Plot of SAXS data as log (intensity) vs. q for a lamellar forming (PI)PS miktoarm star block copolymer.

When analyzing a block copolymer small-angle scattering pattern, for example as shown in Fig. 2, the first step is to compute the series of scattering vector ratios for the Bragg peaks in the pattern. This is done by dividing the scattering vector, q_n, at each peak by the scattering vector q^* of the lowest scattering angle, primary peak. This series of ratios is characteristic of the specific relationship between Bragg planes in a particular type of lattice:

lamellar (one-dimensional lattice):

$$q_n/q* = 1,\ 2,\ 3,\ 4,\ 5,\ 6\ldots$$

simple cubic (sc):

$$q_n/q* = 1,\ \sqrt{2},\ \sqrt{3},\ \sqrt{4},\ \sqrt{5},\ \sqrt{6},\ \sqrt{8},\ \sqrt{9}\ldots$$

body centered cubic (b.c.c.):

$$q_n/q* = 1,\ \sqrt{2},\ \sqrt{3},\ \sqrt{4},\ \sqrt{5},\ \sqrt{6},\ \sqrt{7},\ \sqrt{8}\ldots$$

hexagonal:

$$q_n/q* = 1,\ \sqrt{3},\ \sqrt{4},\ \sqrt{7},\ \sqrt{9},\ \sqrt{11}\ldots$$

cubic OBDD ($Pm\bar{3}n$):

$$q_n/q* = 1,\ \sqrt{(3/2)},\ \sqrt{2},\ \sqrt{3},\ \sqrt{4},\ \sqrt{(9/2)}\ldots$$

cubic gyroid ($Ia\bar{3}d$):

$$q_n/q* = 1,\ \sqrt{(8/6)},\ \sqrt{(14/6)},\ \sqrt{(16/6)},\ \sqrt{(20/6)},\ \sqrt{(22/6)},\ \sqrt{(24/6)}\ldots$$

Often these ratios allow for straightforward identification of different block copolymer morphologies. For instance, the integral peak ratios in Fig. 2 indicate a lamellar morphology. However, a number of complications can arise. A minimum in the form factor scattering which occurs at the same q value as a Bragg peak of the lattice can cause that peak to be attenuated or completely absent in the scattering data. This can confuse the use of the scattering vector ratios to identify the structure. The precise locations of form factor maxima and minima depend on the domain shape (spheres, cylinders, lamellae, bicontinuous struts), on the component volume fractions (i.e., the size of the domains relative to the size of the unit cell), and on the overall scale of the structure. Thus some samples of a given morphology may show all the expected Bragg peaks, while others with slightly different volume fractions may not. Form factor-induced absences in Bragg scattering are a good reason why SAXS data should, as far as possible, be accompanied by TEM imaging. When a scattering pattern is ambiguous due to a peak absence, TEM can provide real space clues as to the identity of the morphology. This then allows the calculation of the form factor scattering, which confirms that the Bragg peak absence results from a form factor minimum. Additionally, broad form factor maxima may appear in the scattering data, and if they are mistaken for Bragg reflections confusion can result.

The preceding discussion has ignored effects such as limited grain size, lattice imperfections, thickness of the interphase between domains, domain size polydispersity, and thermal fluctuations of the lattice. These factors, while not changing the location of Bragg peaks, tend to broaden the scattering maxima, resulting in less distinct patterns. It is possible to calculate the effects of all of these types of structural disorder on the observed scattering intensity profile. However, since different types of structural disorder are superimposed and produce roughly similar peak-broadening effects, and since experimental data is of limited quality, it is generally not practical to try to estimate structural parameters such as average grain size or interfacial thickness by fitting the broadening of Bragg reflections in small-angle scattering data. One can, however, use grain size data and interfacial thicknesses estimated by other techniques (TEM, light scattering, reflectivity, Porod analysis), along with Bragg's law and an appropriate form factor, to simulate scattering data for comparison to experimental results. In this way the plausibility of a given structural interpretation of the data may be demonstrated.

Recently, block copolymer materials have been encountered in which a specific domain shape such as spheres or cylinders is formed, but the domains are not ordered on a lattice. Rather, they form a liquid-like packing, random, yet densely filling space. In this case, the scattering is the product of form factor scattering from the well-defined domain shapes and diffuse, liquid-like interparticle scattering; sharp Bragg reflections are absent. The broad scattering maxima in small-angle scattering data from such a material arise from maxima in the square of the form factor. The TEM images of such materials are difficult to interpret, since the randomly ordered domains produce a confusing projected image. By comparing experimental data to form factor scattering calculated for different domain shapes, it may be possible to determine the morphology.

3. Other Techniques

Although TEM and small-angle scattering are the workhorse techniques for morphological study of block copolymers, other techniques have been utilized. When stained with OsO_4, it has been possible to use field emission gun scanning electron microscopy (FEGSEM) to image diene-containing block copolymer domains at a free surface. In this case, a back scattered electron detector is used to differentiate the stronger back scattered signal from the stained diene domains. The images produced by back scattered FEGSEM have the opposite contrast to TEM images of similarly stained diene-containing block copolymers. In TEM the stained domains are dark, while in FEGSEM they are bright. This technique requires the high-resolution capability of FEGSEM; it cannot be performed with conventional thermionic SEM. One can contemplate using a similar approach to looking at other block copolymer surfaces, so long as one domain can be preferentially stained with a heavy-element-containing compound. It is also possible to use atomic force microscopy (AFM) to observe the morphology of free block copolymer surfaces. The domain structure at the surface provides both topographic contrast and modulated force contrast.

Bibliography

Anderson D M, Bellare J, Hoffman J T, Hoffman D, Gunther J, Thomas E L 1992 Algorithms for the computer simulation of two-dimensional projections from structures determined by dividing surfaces. *J. Coll. Interface Sci.* **148**, 398

Glatter O, Kratky O 1982 *Small Angle X-ray Scattering*. Academic Press, New York

Hirsch P, Howie A, Nicholson R, Pashley D W, Whelan M J 1977 *Electron Microscopy of Thin Crystals*, 2nd edn. Krieger, Malabar, FL

Williams D B, Carter C B 1996 *Transmission Electron Microscopy: Imaging III*. Plenum, New York

S. P. Gido
University of Massachusetts, Amherst,
Massachusetts, USA

Bone Mineralization

Bones can be considered as an assemblage of structural units designed to meet multiple requirements of a specific higher organism. Bones are formed with optimal shapes for enclosing and protecting soft organs and other soft tissues, and for isolating the marrow cavity. The long dimensions, strength, and stiffness of bone are essential mechanical properties for maintaining shape and for locomotion—muscles, which anchor to bones through tendons, move them as levers. The complex nanocomposite material of which bones are formed is also multifunctional. It provides the requirements of stiffness, strength, and elasticity, yet it is readily resorbable, thereby providing important ion storage capacity. Engineering materials designed to have the variety of features found in natural materials, such as bone, might find many applications.

Bone material is a composite material with a complicated arrangement of microscale subunits formed from nanoscale crystals embedded in a biopolymer matrix (Fig. 1). The basic components are calcium phosphate crystals (65 wt.%), collagenous proteins, cells and other macromolecules (lipids, sugars, etc.) (25 wt.%), and water (10%). The unique capabilities of bone derive from a combination of the properties of the crystal/biopolymer material and the layered arrangement of these subunits. Not surprisingly, the synthesis and maintenance of such a complex living material is quite intricate.

Bone formation is accomplished by coordinated multicellular activity. Populations of cells are recruited to a future bony site and are given biochemical signals to produce bone. As will be described in this article, the mineralization of bone is a cell-mediated process involving the specific, highly ordered deposition of a unique form of calcium phosphate confined to precise locations within the organic matrix. Precise assembly of the highly ordered collagenous matrix having specific mineralization sites, and mineral deposition confined to the matrix sites, are essential for bone to meet its multifunctional property requirements. To provide insight into potential biomimetic materials processing strategies, this article presents a review of how bones are formed with a focus on the biology and material science aspects of the mineralization process.

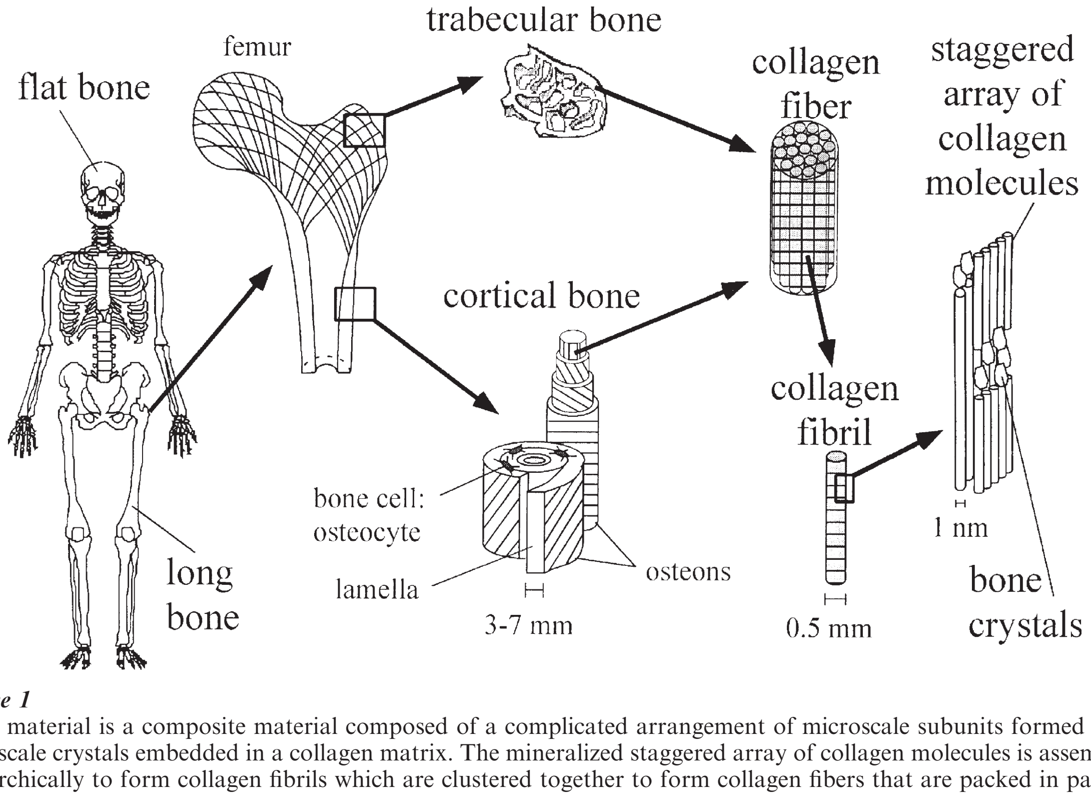


Figure 1
Bone material is a composite material composed of a complicated arrangement of microscale subunits formed from nanoscale crystals embedded in a collagen matrix. The mineralized staggered array of collagen molecules is assembled hierarchically to form collagen fibrils which are clustered together to form collagen fibers that are packed in parallel arrangements to form layers of lamellae. Cortical bone contains osteons that are composed of concentric rings of lamellae that are oriented relative to each other in a twisted plywood fashion. The unique capabilities of bone derive from a combination of the properties of the crystal/collagen material and the layered arrangement of these subunits.

1. Bone Formation and Remodeling

Bone formation, or *osteogenesis,* is initiated during embryonic development by cells that create a nonmineralized fibrous or cartilaginous model in the same shape as the bone to be formed. This model serves as a boundary for skeletal development and provides a scaffold for the skeletal cells to build upon. Cell interactions with the scaffold are very important and play a vital function in cell migration and differentiation. The scaffold is turned into bone by one of two distinct mechanisms: *endochondral ossification* or *intramembraneous ossification* (Marieb 1989, Hall 1994, Bilezikian *et al.* 1996, Raisz and Rodan 1998, Buckwalter *et al.* 1998). The process known as endochondral ossification produces long bones, such as the femur of the thigh, that bear weight and are linked together at joints (Fig. 1). In contrast to long bones, the process termed direct intramembranous ossification forms flat bones, including the bones of the skull (Fig. 1).

1.1 Endochondral Ossification

During embryonic development, precursor chondroprogenitor cells condense in a specific area and become cartilage-forming cells known as *chondrocytes*, which form the cartilage model of the bone (Fig. 2). Other undifferentiated cells in the area surrounding this cartilage zone differentiate into bone-forming cells, called *osteoblasts*, which form a bony collar around the center portion of the cartilaginous long bone model. Osteoblasts are plump cuboidal cells organized in continuous layers adjacent to bony sites that produce *osteoid*, the nonmineralized, collagen-rich bone matrix (Fig. 2). The bony collar starves the cells within and degradation of the cartilage model begins. Blood vessels that have formed within the rigid collar of bone invade the cartilage model bringing resorptive cells and marrow elements.

The blood vessels also provide a conduit for the removal of the degradation products. Eventually the central canal is emptied of cartilage cells and refilled with bone-forming cells and marrow. Osteoblasts within the central canal deposit osteoid around the remaining, very minimal, calcified cartilage to form a network of porous, spongy bone on the cartilage spicules. The bone-covered spicules become the porous *trabecular* bone structures that support the compact bone collar. *Trabecular* bone, also known as *cancellous* bone is composed of small, flat sections of bone (trabeculae) connected by cylindrical struts of bone and is found primarily in the wide ends of long bones (Fig. 1) (Kaplan *et al.* 1994).

The wide ends of long bones are formed and mineralized after the midsection collar of cortical bone is formed. Before the midsection process reaches the ends, secondary centers of ossification are initiated within both ends of the cartilage model. Collar formation and resorption of the central canal occurs in the secondary centers as in the midsection of the bone. However, at the location where the secondary collar meets the primary collar, an active cartilagenous bone growth area remains until skeletal maturity. This

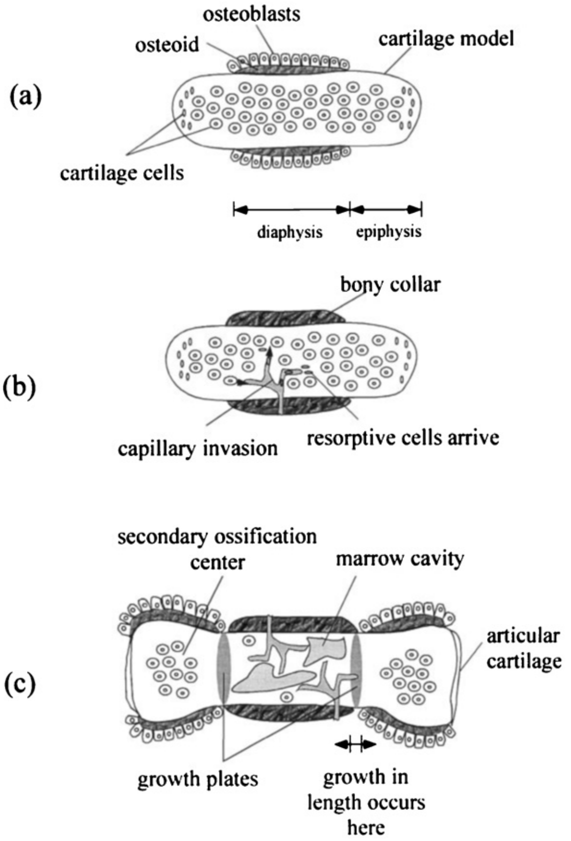


Figure 2
Long bones such as the femur are formed by endochondral bone formation. (a) The cartilage model of the bone is surrounded by *osteoblasts*, which form a bony collar around the center portion (diaphysis) of the cartilage model. (b) The bony collar starves the cells within and degradation of the cartilage model begins. Blood vessels provide a conduit for the removal of the degradation products and marrow influx. (c) Secondary ossification centers initiate at the epiphyses. The junction of the secondary collar and the primary collar (growth plate) remains an active bone growth area until skeletal maturity and allows the bone to grow in length by essentially the same endochondral process. The ends of the bone become covered with articular hyaline cartilage through the localized differentiation and activity of a population of cells within the cartilage model that did not participate in the mineralization process.

area is known as the *epiphyseal growth center* or *growth plate*, and is a three-dimensional zone, which allows the bone to grow in length by essentially the same endochondral process (Buckwalter *et al.* 1998 pp. 183–374). As long bones mature, the distal ends of the bone widen to accommodate future joint stresses. The ends of the bone are finished off with layers of articular hyaline cartilage through the localized differentiation and activity of a population of cells within the cartilage model that did not participate in the mineralization process.

While forming bone, osteoblasts secreting the osteoid bone matrix maintain contact with one another by tentacle-like cytoplasmic projections. Then, as the matrix hardens and the cells become trapped within it, the osteoblasts become the maintenance cells of bone, known as *osteocytes*, which remain connected by extensive cell processes now called canuliculi. It is through the canuliculi that communications regarding ion homeostasis and mechanical strain and stress signals are passed.

1.2 Intramembranous Ossification

In intramembranous bone formation mesenchymal progenitor cells form fibrous connective tissue membranes, which serve as the initial supporting structure or scaffold. Cartilage cells are not a component of intramembranous bone formation. The transformation of the initially formed organic scaffold involves the secretion of unmineralized osteoid matrix around the fibrils of the connective tissue membranes by osteoblasts. The accumulations of osteoid eventually are mineralized and become a network of porous, trabecular bone.

A multilayered connective tissue, the *periosteum*, which contains fibroblasts, loose connective tissue with nerve fibers, blood vessels, and osteoblasts, forms to enclose the trabecular bone. With time, new bone fills the spaces beneath the periosteum and within the outer portion of the trabecular bone, thereby forming a compact bone shell that encloses the inner remaining trabecular bone in a sandwich construction. The shape of the bone remains that of the initial fibrous tissue model.

1.3 Remodeling

Immediately following initial bone formation by either endochondral or intramembranous ossification mechanisms a process known as *remodeling* commences performed by osteoclasts (Bilezikian *et al.* 1996, Kaplan *et al.* 1994). Osteoclasts are derived from a macrophage–monocyte lineage and are the major resorptive cells of bone. They are large (20–100 microns in diameter) and have multiple nuclei. Osteoclasts bind to the bone surface through cell attachment proteins called integrins and resorb bone by first lowering the pH of the local environment below them by a release of H^+ protons. The low pH dissolves the bone crystals, and the organic components of the matrix are removed by acidic proteolytic digestion. Osteoblasts reform the bone but with a new microstructure. Unlike the relatively unorganized woven structure of rapidly formed first bone, the bone formed after remodeling is precisely ordered lamellar bone. Lamellar bone consists of layered structures at many length scales as shown schematically in Fig. 1.

Additional remodeling occurs during skeletal maturation to effect growth in diameter of long bones. Growth in diameter of long bones occurs by a process known as *appositional growth*. A new deposition of bone by the direct intermembraneous route occurs on the outer surfaces of the shafts with concurrent resorption of the interior by osteoclasts, which results in a growth in bone diameter without a thickening of the cortical bone. The growth in length is accomplished through the growth plate.

Even following skeletal maturity, bone continues to remodel throughout life to adapt its material properties to best meet the mechanical and metabolic demands placed on it. All bone surfaces are lined by resting osteoblasts and undifferentiated stem cells that respond to biomechanical and chemical stimuli. The bone matrix probably influences cell behavior by deforming under mechanical loads and transmitting this information to the cells. Both osteoblasts and osteoclasts and osteocytes within the bone can respond by increasing the rates of bone formation or resorption, respectively. The communication between osteoblasts and osteoclasts controls the net turnover of bone. When osteoblasts are stimulated by the biochemical or mechanical signals they receive, they communicate to osteoclasts to modulate the rate of bone resorption to maintain the total bone mass.

Normal, healthy remodeling occurs under close biochemical control and results in no net change in bone mass. Osteoporosis is an example of a diseased state of bone, inevitable in the elderly, where osteoclastic bone resorption exceeds that of osteoblast bone formation resulting in a net loss of bone. Trabecular bone is mainly lost, and unfortunately cannot be regenerated after a certain stage of connectivity is lost (Kaplan *et al.* 1994).

2. The Mineralization of Bone

In summary, mineralization occurs by: (i) primary heterogeneous nucleation within the hole zones of collagen fibrils; (ii) crystal growth within and beyond the hole zones; and (iii) secondary nucleation induced by previously formed crystals to reach a final mineral density of 65 wt.%, 25 wt.% organic matrix, and 10 wt.% water.

2.1 Collagen Matrix Organization

Bone mineralization, the process by which the organic bone matrix becomes filled with calcium phosphate nanocrystals, occurs in a specific, highly ordered process. The process is mediated by osteoblasts and confined to the organic osteoid matrix produced by osteoblasts. Osteoid is primarily composed of Type I collagen fibrils arranged as shown in Fig. 1 to create the macro- and microstructures of bone. Each collagen molecule is constructed of three polypeptide chains composed of approximately 1000 amino acids each that are helically wound around each other and cross-linked by hydrogen and covalent bonds, resulting in a fairly rigid, rod-like molecule approximately 300 nm long (Fig. 3). These molecules are assembled into fibrillar structures by a process of axial and radial aggregation, which results in a specific tertiary structure having a 67 nm axial periodicity and 40 nm gaps or hole zones between the ends of the collagen molecules (Fig. 3) (Glimcher 1998). The hole zones align in channels through the fibrils.

2.2 Initiation of Mineralization

While several theories on the initiation of mineralization have been proposed after years of intensive study, electron microscopy studies on nonaqueously prepared specimens have repeatedly confirmed that initial deposition of the bone crystals occurs primarily within the collagen fibril hole zones shown in Fig. 3 (Glimcher 1998). The periodic deposition of mineral crystals within the collagen fibrils is demonstrated in the electron micrograph of tendon (Fig. 4), which has a similar periodic structure to bone, but is more easily visualized because it is less dense and has a more aligned organic matrix. Studies of fish bone have also conclusively demonstrated that initial mineralization occurs within the collagen hole zones (Glimcher 1998). While there are some similarities between bone and tooth mineralization, those readers interested in enamel and dentine mineralization can find specific information elsewhere (Robinson *et al.* 1995).

Bone mineralization is thought to commence heterogeneously within the supersaturated hole zones. However, the small size of the hole zones precludes actual testing of the microchemistry and protein content within them. A schematic of what is believed to occur chemically within a hole zone at the moment of mineralization is shown in Fig. 3 (Glimcher 1998). Ions of calcium and phosphate are shown loose and bound to localized matrix proteins arranged with a molecular periodicity that may serve to nucleate the mineral phase heterogeneously. It is known that initiation of mineralization requires a combination of

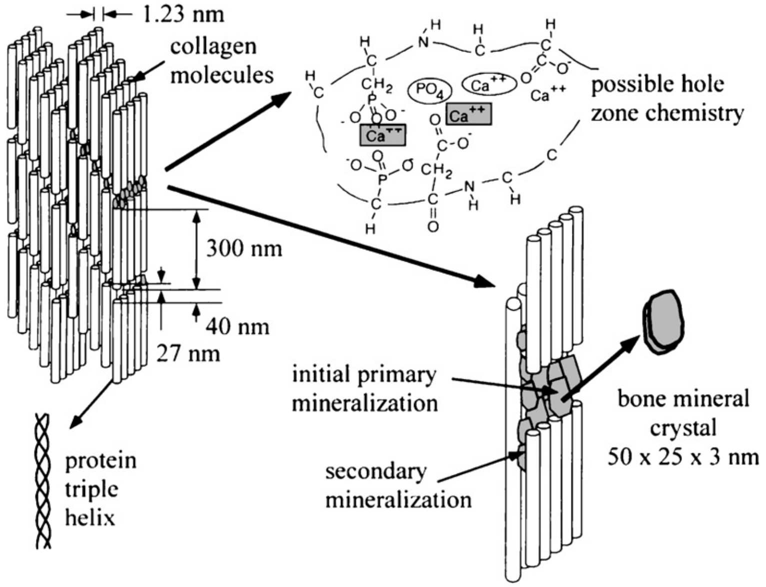

1.23 nm
collagen molecules
possible hole zone chemistry
PO4
Ca++
300 nm
40 nm
27 nm
initial primary mineralization
secondary mineralization
bone mineral crystal 50 x 25 x 3 nm
protein triple helix

Figure 3
An illustration of the arrangement of collagen molecules within a collagen fibril with a 67 nm periodic structure that provides hole zones for nucleation and growth of the bone mineral. The schematic of the chemistry within the microenvironment of a collagen hole zone illustrates the probable nucleation processes involving free and bound mineral components and bound side chain groups with a molecular periodicity that serves to heterogeneously nucleate the mineral component.

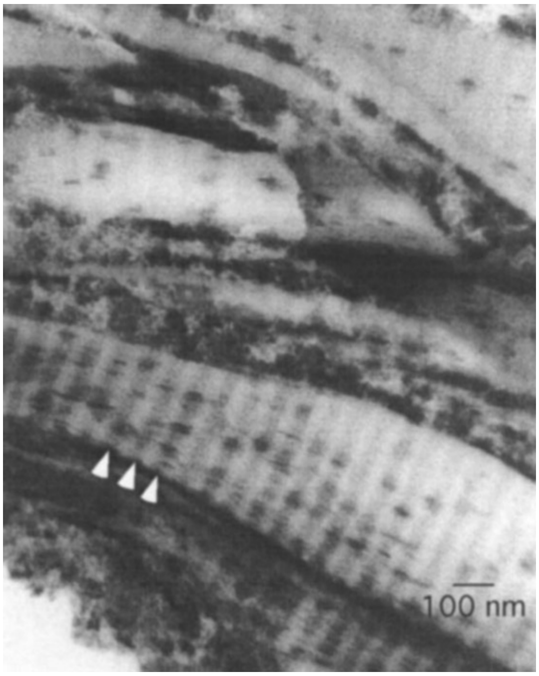


Figure 4
A region of calcified turkey leg tendon showing the highly ordered, repeating pattern of mineral deposition with a collagen fibril (micrograph kindly supplied by Dr W J Landis).

events, including increases in the local concentration of precipitating ions, formation or exposure of mineral nucleators, and removal or modification of mineralization inhibitors. It is the osteoblasts located adjacent to the osteoid which take ions from the capillaries and directionally secrete the mineral components into the matrix, where mineralization occurs.

Mineralization does not occur on the cells because the mineralization is governed by heterogeneous nucleation factors. In addition to collagen, the organic matrix contains proteoglycans and noncollagenous proteins including growth factors, cytokines, bone inductive proteins, osteocalcin, osteopontin, bone sialaprotein, and glycosolated phosphoproteins, which govern and regulate various aspects of mineralization (Kaplan *et al.* 1994, Bilezikian *et al.* 1996). Many of the phosphoproteins are believed to be localized in the hole zones of the collagen fibrils and participate directly in mineralization. These bound proteins are phosphorylated enzymically prior to mineralization and are therefore postulated to assist in the heterogeneous nucleation of bone crystals.

While the phosphoproteins play a positive role in matrix mineralization, proteoglycans appear to inhibit mineralization, and are theorized, for example, to prevent the mineralization of the cartilage on articular joint surfaces. Proteoglycans consist of a protein core and one or more glycosaminoglycans, chains of negatively charged polysaccharides that vary considerably in size and number. The high density of negative charges from the side chains are postulated to bind calcium, repel phosphate, and sequester mineral clusters, thereby preventing mineral calcification (Buckwalter *et al.* 1998).

In very early embryonic bone, initial mineral deposits also occur to a small extent in cell structures called vesicles, which are small extracellular membrane compartments pinched off from a cellular membrane protrusion. The concurrent appearance of the vesicles with the initiation of mineral deposition has led many to postulate that the matrix vesicles were mineralization catalysts (Landis 1986). However, matrix vesicles are not required for mineralization of bone as evidenced by the vesicle-free rapid remodeling of all first formed embryonic bone that occurs within hours of initial mineralization and the mineralization of many species of fish bone (Glimcher 1998). New bone tissue is routinely synthesized during embryogenesis and postnatally at a time when matrix vesicles are absent or near absent from the tissue.

2.3 Primary and Secondary Nucleation

Further mineralization of the collagen occurs by additional primary nucleation and secondary nucleation by the crystals already formed within the hole zone. Eventually after the available hole zone is filled, crystals grow and extend beyond the hole zones into the pore space between the molecules of the collagen fibrils (Fig. 3). During the early stages of mineralization, the mineral phase has an axial periodicity of ~700 Å, which reflects the initial mineralization position within the hole zones of the collagen fibrils that also have an axial repeat of ~700 Å. When mineralization of the collagen is completed, this periodicity is lost due to the deposition of crystals throughout the available spaces. The secondary growth of crystals into the interfibrillar space greatly stiffens the bone.

2.4 Bone Inductive Factors

Bone inductive factors (e.g., bone morphogenetic proteins (BMPs)) and other cytokines found localized in skeletal tissues are not directly involved in the physics of nucleation of bone mineral, but function indirectly as potent cell-signaling molecules. These molecules regulate cell proliferation, differentiation, migration, and activity (Bilezikian *et al.* 1996, Hall 1994, Buckwalter *et al.* 1998, pp. 87–98). BMPs belong to a large family of secreted signaling molecules originally named for their remarkable ability to induce the formation of new cartilage and bone when implanted at non-bony sites. Since the first discovery of BMPs as a component of demineralized bone matrix, a material that is routinely used to heal bone defects, a large family of related proteins has been defined that appear to control many developmental events in different organisms (Buckwalter *et al.* 1998).

Structurally, BMPs are classified as belonging to the transforming growth factor beta (TGF-β) superfamily that mediates key events in growth and development. Several BMPs are expressed in bone tissue implying multiple BMPs are involved in mediating skeletal patterning and bone cell differentiation. Several BMP products are being developed for clinical applications; for example, to stimulate either bone formation (fracture repair, dental extraction, craniofacial reconstruction, and spinal fusions) or cartilage or tendon and ligament formation. The BMPs must be used with a scaffold material of some kind in order to generate tissue. Skeletal cells are also added to improve the regeneration rate. The design, specification and fabrication of cells, biomaterials and biomolecules which directly replace or stimulate the organism to synthesize specific tissues or organs is known as tissue engineering.

3. *The Chemical and Physical Structure of Bone Mineral*

It has been known since 1926 that bone mineral is calcium phosphate, with an apatite-like structure containing carbonate. However, more than 75 years later, scientists are still striving to identify the chemical and physical structure of bone mineral conclusively (Glimcher 1998). The characterization of the exact size, shape, and chemical composition of bone crystals is hampered by the poorly crystalline nature of bone mineral and the organic component. Both aspects contribute to a reduction or broadening of the signals emitted from the mineral phase during the various characterization methods that have been applied.

All calcium phosphate (CaP) apatites have the same formula $Ca_5(PO_4)_3X$ where X can be for example an OH^- ion (hydroxyapatite) or a CO_3^{2-} ion (carbonated apatite). The basic CaP apatite structure is hexagonal with space group $P6_3/m$ and approximate lattice parameters $a = 9.4$ and $c = 6.9\,\text{Å}$ with two formula units per cell. The apatite structure is very tolerant of ionic substitutions. It is therefore not surprising that bone mineral formed in the presence of a bloodstream containing a variety of metals and ions contains small, but significant, amounts of impurities such as HPO_4 ion, CO_3 ion, Al, Pb, citrate, and others whose positions and configuration are not completely known. Na^+, Mg^{2+}, and K^+ may also be associated with bone apatite as surface-adsorbed ions. Chemical analysis has shown that bone mineral contains about 5% CO_3 and about 5–10% HPO_4. With time the Ca/P ratio increases to values approaching that of pure apatite (1.67), however in general bone apatite is calcium deficient.

X-ray diffraction spectra of bone apatite show only broad overlapping peaks that correspond to those of apatite. No crystalline phases other than apatite have been detected in any bone samples that were prepared non-aqueously. The poorly crystalline, nanocrystalline nature of the mineral makes it very susceptible to alteration during specimen preparation involving water or heat. Hence the literature contains a sizeable portion of articles not actually reporting on bone mineral, but on some altered form of bone mineral. For example, several precursor phases have been reported to exist within bone samples that were prepared with water exposure or high temperatures (Termine 1972). Later studies have disproved these hypotheses (Glimcher 1998). The high initial content of organic matrix relative to bone mineral led several investigators to believe that bone mineral was first deposited as an amorphous calcium phosphate because the x-ray diffraction pattern was so obscured by the organic content. Studies that use a low-temperature, nonaqueous technique to obtain organic-free crystals of bone have shown that even the very earliest mineral deposited in bone and cell culture matrices is carbonated apatite of the same basic structure and chemistry as the mineral in mature bone (Glimcher 1998).

Electron micrographic observations of isolated bone crystals and tomographic three-dimensional computer reconstruction analyses of calcified and ossified turkey tendon have vividly and conclusively established that bone crystals exist as small, thin plates with approximate dimensions of 35–100 nm × 50–25 nm × 1–3 nm (Glimcher 1998). There is a clear variability in the size distribution of the plate-like crystals that is a function of species. For example, mineral crystal size decreases in the following order: herring bone crystals >cow bone crystals >chicken bone crystals >mouse bone crystals. A sample of crystals isolated from the organic matrix of cow bone is shown in Fig. 5. No rod- or needle-like crystals have been obtained in the crystals isolated from the organic phase, however the plate-like crystals appear needle-like when viewed on edge within a section of bone and have resulted in several inaccurate reports of needle-like bone crystals in the literature.

The structural aspects of bone mineral are best determined by the vibrational spectroscopies, which have been extensively used to determine the location of the carbonate and HPO_4 ions and the surprising absence of OH groups. From resolution enhanced FTIR studies the mineral crystals of bone have been determined to be Type AB carbonated apatite without detectable OH groups (Glimcher 1998). While the x-ray diffraction pattern of bone apatite is nearly identical to that of hydroxyapatite there are several chemical and structural differences between bone apatite and synthetic hydroxyapatites containing carbonate. The most obvious features are nonapatitic carbonate positions, in addition to crystalline Type A and B carbonate locations, and non-apatitic phosphates.

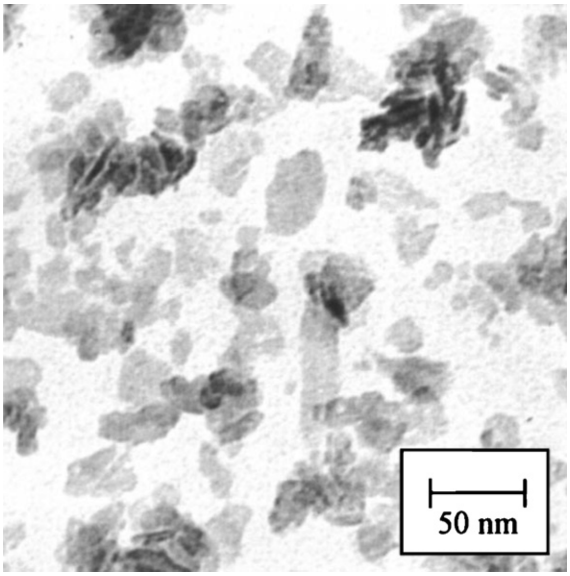


Figure 5
An electron micrograph of dispersed bone crystals isolated from compact bovine bone by non-aqueous processing. The crystals are irregular plates with approximate dimensions of 35 nm × 25 nm × 3 nm.

^{31}P NMR spectroscopy has shown that the HPO_4 ion has a unique configuration in bone quite distinct from all other CaP phases which contain HPO_4 ions such as brushite, monetite, and octacalcium phosphate. However, the PO_4 spectra are conclusively apatite although some of the PO_4 in very young bone is in an unstable environment. Most importantly, a significant chemical difference between bone apatite and other CaPs is the near-absence or absence of hydroxyl groups in the bone apatite which been proven repeatedly by chemical methods and FTIR and NMR spectroscopy (Glimcher 1998). Calcium deficiencies and carbonate ion substitutions are necessary to maintain charge neutrality.

The complex crystal chemistry and structure of bone apatite is not very stable and has been observed to change with time *in vivo* during a process known as maturation. Maturation is manifested through changes in the short-range order of the crystals and the concentration and location of highly labile, reactive CO_3 and HPO_4 groups. Distinctive markers of bone mineral maturation are an increase with time of stable carbonate and a decrease of labile CO_3 and HPO_4 ions. Mature bone crystals are more crystalline, more stable and less reactive than young, newly formed crystals.

The unique chemical and physical features of bone apatite impart important biological properties that are critical to normal bone function, including the ion substitutions that increase solubility to increase the accessibility of the ions stored in bone mineral. Vital ion storage of nearly 99% of the body calcium, about 90% of the body phosphorus, and from 90% to 50%, respectively, of the total body sodium and magnesium are associated with bone mineral. The small size and large surface area of the crystals permit very rapid and extensive exchange, addition and dissolution of the mineral mass and the constituent ions and atoms in the lattice. The highly ordered location and orientation of the very small nanocrystals located within the collagen fibrils also contribute significantly to the structural rigidity and strength of bone while allowing for flexibility without fracture of the bone.

4. Summary

Bone formation involves the coordination of a variety of cellular activities in specific locations. Multifunctional property requirements are met by a tailored formation process and variation of material architecture among bones and even within a single bone. Skeletal cells build a mineralized tissue through the use of an organic scaffold and maintain it through remodeling during the lifetime of the organism. The cell activities and the mineralization of the organic matrix are governed by organic macromolecules found both locally and systemically. The mineralization is confined to specific locations within the collagen matrix and the mineral formed has a unique chemical composition and physical structure. The complexities of bone cell microenvironments cannot be reproduced *in vitro*, however recent advances in cellular and molecular biology have made it possible to use cell signaling molecules and/or cells in combination with an inert porous scaffold material to regenerate bone tissue *in vivo*. The fact that a simple mixture of collagen fibrils with precipitated apatite prepared *in vitro* does not produce a substance with mechanical properties even close to that of bone, demonstrates that the hierarchical arrangement and precise cell-mediated construction of bone is of the utmost importance.

Bibliography

Bilezikian J P, Raisz L G, Rodan G A (eds.) 1996 *Principles of Bone Biology*. Academic Press, San Diego, CA

Buckwalter J A, Ehrlich M G, Sandell L J, Trippel S B (eds.) 1998 *Skeletal Growth and Development: Clinical Issues and Basic Science Advances*. American Academy of Orthopedic Surgeons, Rosemont, IL

Glimcher M J 1998 The nature of the mineral phase in bone: biological and clinical implications. In: Avioli L V, Krane S M (eds.) *Metabolic Bone Disease and Clinically Related Disorders*. Academic Press, San Diego, CA, pp. 23–50

Hall B K (ed.) 1994 *Bone*. CRC Press, Boca Raton, FL, Vols. 1–9

Kaplan F S, Hayes W C, Keaveny T M, Boskey A, Einhorn T A, Iannotti J P 1994 Form and function of bone. In: Simon

S R (ed.) *Orthopedic Basic Science*. American Academy of Orthopedic Surgeons, Rosemont, IL, pp. 127–84

Landis W J, Paine M C, Hodgens K J, Glimcher M J 1986 Matrix vesicles in embryonic chick bone: considerations of their identification, number, distribution, and possible effects on calcification of extracellular matrices. *J. Ultrastruct. Mol. Struct. Res.* **95**, 142–63

Marieb E N 1989 Bones and bone tissue. *Human Anatomy and Physiology*. Benjamin/Cummings, Chap. 6, pp. 152–66

Raisz L G, Rodan G A 1998 Embryology and cellular biology of bone. In: Avioli L V, Krane S M (eds.) *Metabolic Bone Disease and Clinically Related Disorders*. Academic Press, San Diego, CA, pp. 23–50

Robinson C, Kirkham J, Shore R (eds.) 1995 *Dental Enamel Formation to Destruction*. CRC Press, Boca Raton, FL

Termine J D 1972 Mineral chemistry and skeletal biology. *Physiol. Rev.* **49**, 760–91

L. T. Kuhn
Children's Hospital and Harvard Medical School,
Boston, Massachusetts, USA

C

Calamitic Liquid Crystals

The current state of the art can be found in the handbooks by Demus *et al.* (1998) and Collings and Patel (1997). Many details concerning phase types and structures are presented in Demus and Richter (1978), Gray and Goodby (1984), Goodby (1998), and Seddon (1998).

1. Position of Liquid Crystals Between the Solid State and the Isotropic Liquid State

In solid molecular crystals there is three-dimensional long-range order of the positions and long-range order of the orientations of the molecules. The molecules exist in one or just a few preferred conformations. In ordinary materials, at the melting temperature the solid crystal transforms to the isotropic liquid, which has random (or short-range) order of the positions and orientations of the molecules. If the molecules are flexible, they exist in a large number of different conformations following a Boltzmann distribution.

Liquid crystals, also called mesophases, are intermediate between these two states. That is, the positions of the molecules possess less than three-dimensional long-range order, the orientations may be short- or long-range ordered, and the molecules (when they are flexible) can exist in several conformations. The molecules have several degrees of freedom of mobility, that is, rotations around the long and short molecular axes, fluctuations around the direction of the orientation, and relatively high translational mobility. Therefore liquid crystals possess weak elastic constants and are liquid-like as far as their viscous properties are concerned.

There are structures with three-dimensional order, in which the globular molecules have rotational freedom. These are called "plastic crystals" and have cubic structures.

Also in the case of non-globular or elongated molecules degrees of rotational freedom about a preferred axis may be excited in crystals while still retaining three-dimensional positional order. Such structures are called "hexagonal plastic crystals"; the structures formerly called "ordered smectics" belong to this category. Since the latter are closely related to liquid crystals, they are designated as "soft crystals."

In crystals built up of molecules with flexible groups, the latter may exist in several conformations, causing some disorder in the crystals, termed "condis crystals" (*con*formationally *dis*ordered crystals). Condis crystals often occur with mesogenic molecules. It should be mentioned that in some cases the exact distinction between liquid crystals and solid crystals is difficult.

2. Molecular Structure and Phase Structure

Molecules forming liquid crystals can have quite different chemical structures and different shapes. Table 1 shows some representative examples. Among the classical rod-like (calamitic) molecules liquid crystals were discovered in the late nineteenth century. They consist of more or less stiff moieties, different in chemical nature, but providing the necessary elongated shape. Because in most cases the molecules have rotational freedom around their long axes, they are represented symbolically by rotational cylinders. On the other hand, disk-like molecules that form liquid crystals have been discovered more recently (see Chandrasekhar 1998). Intermediate mesogenic molecular structures with branches or lathlike shape ("sanidic liquid crystals"), dimers, oligomers, and polymers have also been found.

The present chapter is devoted to calamitic liquid crystals. In the liquid crystalline state the rod-like molecules are generally oriented with their long axes parallel to one another. Obeying this condition, there are many possible ways of packing the molecules, and a large number of phase structures are possible.

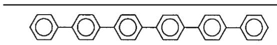
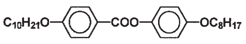
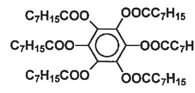
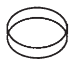
3. Classification of Liquid Crystals

There are about 80000 compounds with liquid crystalline phases. This vast number of phases can be reduced to a limited number of phase types.

In early studies the phase types were classified into nematic, smectic, and cholesteric phases. At present more than 20 different smectic phase types are known.

The basis of modern classification initially has been the selective miscibility of liquid crystalline phases (Sackmann and Demus 1966, 1973). That is, completely miscible phases were labeled as belonging to

Table 1
Molecular structures.

Rod-like = calamite	Disk-like = discotic
 	
symbol:	

the same phase type. Incomplete miscibility does not allow a classification. The phase types obtained by this procedure showed typical textures when seen under the microscope using polarized light. In fortunate cases the classification is possible simply on the basis of texture observation only, and in fact this is the most widespread method for a first characterization of liquid crystals after synthesis.

After applying x-ray methods more extensively to liquid crystals, it was found that the phase classification based on miscibility and textures corresponds to the classification according to structural features. In all more complicated cases, methods allowing the characterization of phase structures (x-ray, electron and neutron scattering, NMR, ESR, and other spectroscopic methods) are applied. These methods also allow the detection of subtypes of the classical phase types.

The general phase types of calamitic liquid crystals are:

- *nematic*: orientational order of the molecular long axes
- *smectic*: layer structures, several possibilities of the state of order inside the layers and several possibilities of mutual arrangement of the layers, orientational order, and eventual positional order
- *soft crystals*: formerly called "ordered smectics," layer structures with three-dimensional orientational and positional order, lower order than ordinary crystals
- *cubic*: structures with micellar lattice units or complicated interwoven networks, three- dimensional order of the lattice units, liquid-like on the molecular level.

4. *Phase Types of Non-chiral Compounds*

The main phase types of non-chiral compounds are compiled in Table 2 and displayed in Fig. 1.

4.1 Nematic Liquid Crystals

In the nematic (N) phase the molecules are oriented parallel one to another (Fig. 1). They have translational freedom, and rotation about the long and short axes is possible. When the molecules are flexible, they exist in all possible conformers in a Boltzmann distribution. The main difference between the nematic and ordinary isotropic liquids is the preferred orientation of the molecular long axes (director of the N phase).

In many compounds nematic/smectic phase transitions occur. In the weakly first-order nature of these phase transitions there are strong fluctuations near the transitions, which can be observed in N phases as "cybotactic groups," that is, microscopic clusters of smectic layers within the nematic structure. The cybotactic groups substantially influence the properties of nematic phases.

Because their structure is closely related to those of isotropic liquids, nematics have physical properties similar to them, especially the low viscosity. In contrast they possess optical birefringence, anisotropy of

Table 2
Main phase types of non-chiral calamitic liquid crystal compounds.

Phase type	Positional order		Bond orientational order	Molecular orientation
	Within planes	Between planes		
Nematic				
N	–	–	s	
Smectics				
A	s	s	s	orthogonal
C	s	s	s	tilted
O (C_{alt})	s	s	s	alternating tilt
M	s, hexatic	s	l	tilted
B (hexatic)	s, hexatic	s	l	orthogonal
I	s, hexatic	s	l	tilt to apex of hexagon
F	s, hexatic	s	l	tilt to side of hexagon
Soft crystals				
L (B cryst)	l, hexagonal	l	l	orthogonal
J (G′)	l, pseudo-hexagonal	l	l	tilt to apex of hexagon
G	l, pseudo-hexagonal	l	l	tilt to side of hexagon
E	l, orthorhombic	l	l	orthogonal
K (H′)	l, monoclinic	l	l	tilt to side a
H	l, monoclinic	l	l	tilt to side b

s = short-range order; l = long-range order.

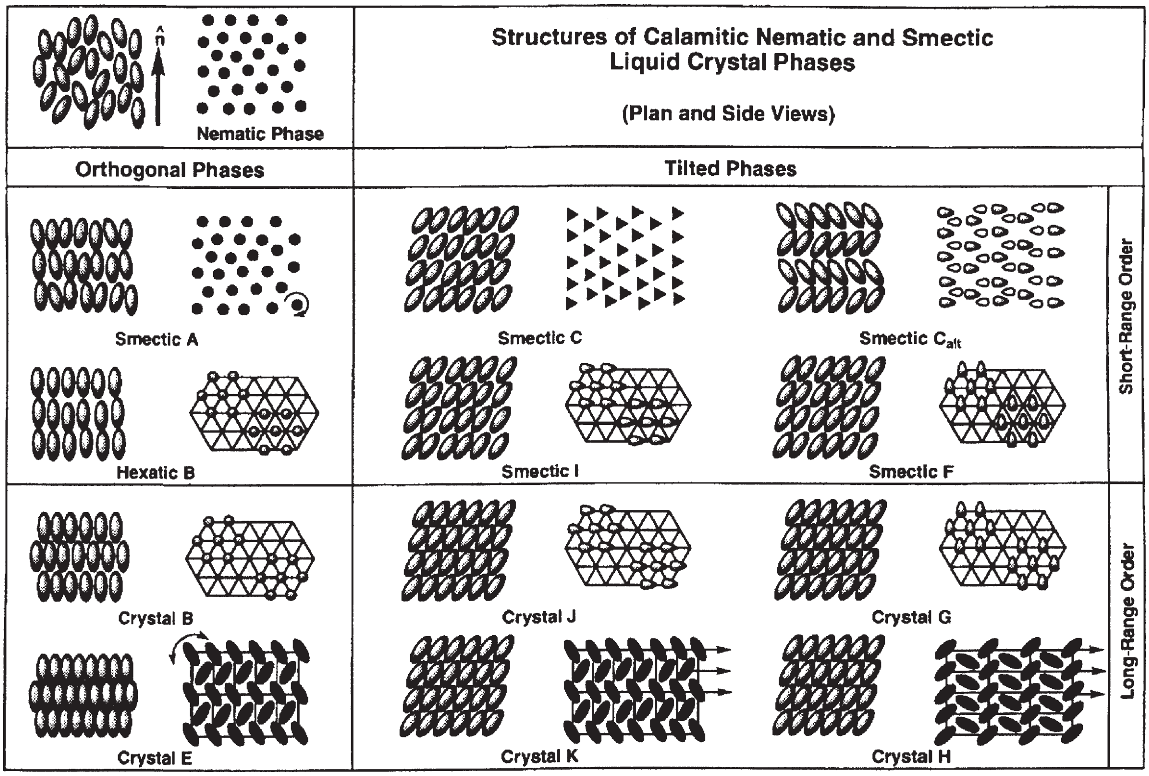

Nematic Phase
n̂
Structures of Calamitic Nematic and Smectic Liquid Crystal Phases
(Plan and Side Views)
Orthogonal Phases
Tilted Phases
Smectic A
Smectic C
Smectic C$_{alt}$
Short-Range Order
Hexatic B
Smectic I
Smectic F
Crystal B
Crystal J
Crystal G
Long-Range Order
Crystal E
Crystal K
Crystal H

Figure 1
Main phase types of calamitic compounds (reproduced by permission of Wiley-VCH from 'Handbook of Liquid Crystals', 1998).

the electric and magnetic suceptibility, and anisotropic elastic properties. The combination of anisotropy and fluidity makes them useful as materials for electrooptic displays.

4.2 Smectic Liquid Crystals

In general, smectic (Sm) liquid crystals have layer structures.

Within the layers Sm A phases have liquid-like (random) order of the molecular centers with the molecules oriented orthogonal with respect to the layer planes (see Fig. 1). The molecules have high mobility, for example rotation about long and short axes, and relatively high translational freedom. The rotation around the long axes, which is hindered by only a very small potential, averages their properties in the lateral directions and causes uniaxial physical anisotropic properties. As exceptional cases, biaxial Sm A phases have recently been detected.

Sm C phases can be considered as tilted variants of the Sm A phases (Fig. 1). Due to the tilt they are biaxial, and they show characteristic optical textures different to those of A phases. There is one variant Sm C$_{alt}$ (also called Sm O phase), in which the tilt directions of adjacent layers alternate. Because the molecules possess some polarity, this phase is an example of an antiferroelectric structure.

In the group of the "hexatic" phases Sm B, I, F, M, the local cross-sectional packing of the molecules within the layers is hexagonal and has a correlation length of about 15–60 nm, indicating the relatively large lateral positional order within the layers, which, however, formally is short-range. The coherence between adjacent layers is weak and short-range. In the Sm B phases the molecules are oriented perpendicularly to the layer planes (Fig. 1). In the F and I phases the molecules are tilted (Fig. 1), in a direction towards the side of the hexagon in the F phases and towards the apex of the hexagon in the I phases. The Sm M phases also belong to this group of hexatic phases; however, details of their structure are not yet known. Because the molecules are not biased in their rotation around the long axes, the effective molecular shapes can be described by rotational cylinders.

Figure 1 shows the most commonly occurring "normal" structures of the respective phase types. In fact especially in Sm A and C there are additional variants, explained in Fig. 2. In the intercalated

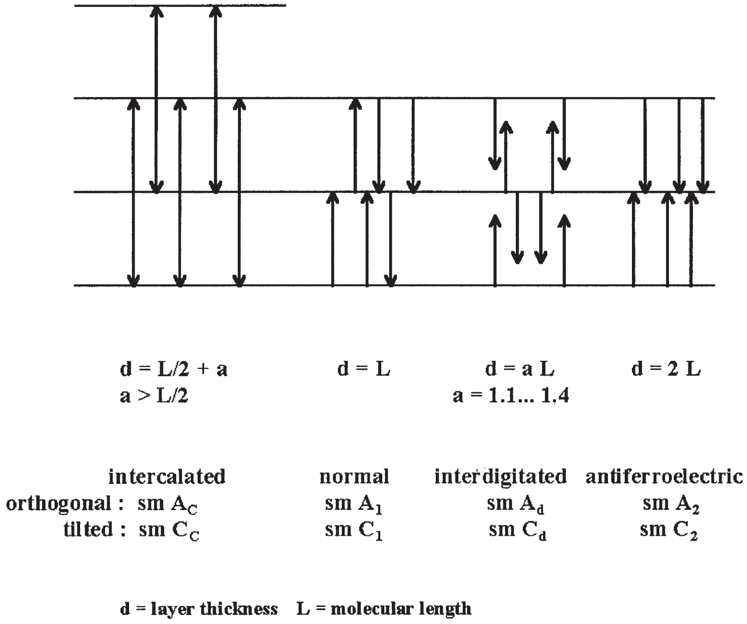


Figure 2
Subtypes of smectic A and C phases.

phases the molecules of adjacent layers overlap one another up to half of their length *L*. Intercalated phases have been found in compounds with lateral aromatic branches, diols, and dimers. The interdigitated phases frequently occur in compounds with endstanding polar groups (cyano, nitro, and others) due to association into dimers. The polar compounds also can form double layers having double the molecular length. The polar groups in adjacent layers are oriented antiparallel, producing a structure with antiferroelectric character. The corresponding phases with orthogonally oriented molecules are subtypes of SmA phases, those with tilted molecules are subtypes of SmC phases.

Among the Sm A_2 and C_2 phases modulated phases (also called crenellated phases) may occur, that is, blocks of the double layers are shifted in longitudinal direction by $L/2$ and modulate the phase to give a ribbon structure.

In twist grain boundary (TGB) phases, which occur in Sm A and C, blocks of the layer structure are turned around by a certain angle to give helical dislocations. TGB phases have frequently been found in the phases of chiral compounds.

4.3 Soft Crystal Phases

When there is long-range correlation between the layers with ordered molecules, the structures have three-dimensional order. There are several phases of this type (Table 2 and Fig. 1) which have certain similarity to Sm B, F, and I phases, and were formerly called "ordered smectic phases." Because of their three-dimensional state of order, in the actual nomenclature they are designated "soft crystal phases." From the chemical standpoint, they occur in calamitic compounds as "low-temperature" phases, relative to the nematic and smectic phases.

The soft crystal B phases (also called L phases) are three-dimensional smectic B phases, sometimes existing in the same compound and divided by a phase transition with quite small transition effects. There are several variants with different stacking sequences of the layers. The soft crystal G and J phases are the three-dimensional variants of the smectic F and I phases.

In addition to these soft crystals, which are related to the hexatic smectics, there are phases with "herringbone" order within the layers, which is due to a biasing of the rotation of the molecules around their long axes. In the soft crystal E the molecules are orthogonal to the layers. The molecules within the layers are arranged in a monoclinic packing, with a zig-zag orientation of the preferred lateral molecular axes, looking from above like a herringbone. The soft crystal K and H phases possess herringbone order within the layers and the molecules are tilted towards the different sides of the monoclinic unit cell.

The soft crystal phases are usually investigated within the area of liquid crystals, because the high mobility and the chemical nature of the molecules, the defect and layer structure of the soft crystals, and the small transition effects between them and real liquid crystals point to many similarities with liquid crystals.

It should be mentioned that in single compounds quite a lot of phases have been found whose structural features cannot be clarified in detail, and whose classification into one of the known phase types remains open. In several cases the features point to phases with structures different from the known phase types.

5. *Phase Types of Chiral Calamitic Molecules*

All the above mentioned phase types occurring in non-chiral compounds exist as chiral variants in compounds with chiral molecules (Table 3). They have the same basic structures, but most of them in addition possess a helical arrangement, which is the reason for their extraordinary optical and electrooptical properties. The chiral character of the phases is indicated by a star in the short designation, e.g. N*, Sm C*.

Despite differences in the phase structures, phases of the corresponding types of non-chiral and chiral compounds show complete miscibility.

5.1 Chiral Nematic Phase

Figure 3(a) displays a sketch of the chiral nematic phase, often called ‘‘cholesteric’’ phase because it was discovered first in cholesterol derivatives. The quasi-layers have nematic character, but the preferred directions of adjacent layers are turned around a small angle producing a helical or screw-like structure. This helical structure has a characteristic pitch *p* with a magnitude of a few micrometers in most cases. Depending on the sense of the chirality of the molecules, the structure corresponds to a right or left screw in character. Racemic mixtures of the chiral molecules produce ordinary nematic phases.

The helicity of the structure is the reason for the selective reflection of circularly polarized light, strong optical activity, and special textures of chiral nematic

Table 3
Main phase types of chiral calamitic liquid crystal compounds.

	Positional order		Bond orientational order	Molecular orientation	
Phase type	Within planes	Between planes			
Nematic					
N*	—	—	s		helix
Smectics					
A*	s	s	s	orthogonal	no helix, electroclinic
C*	s	s	s	tilted	helix, ferroelectric
C^*_A	s	s	s	alternating tilt	helix, antiferroelectric
C^*_α, C^*_γ	s	s	s	alternating tilt	helix, ferrielectric
B*	s, hexatic	s	l	orthogonal	no helix
I*	s, hexatic	s	l	tilted to apex of hexagon	helix, ferroelectric
I^*_A	s, hexatic	s	l	alternating tilt to apex of hexagon,	helix, antiferroelectric
F*	s, hexatic	s	l	tilted to face of hexagon	helix, ferroelectric
F^*_A	s, hexatic	s	l	alternating tilt to face of hexagon,	helix, antiferroelectric
Soft crystals					
B*	l, hexagonal	l	l	orthogonal	no helix
J*	l, hexagonal	l	l	tilted to apex of hexagon	no helix, ferroelectric
G*	l, hexagonal	l	l	tilted to face of hexagon	no helix, ferroelectric
K*	l, monoclinic	l	l	tilted to side a	no helix
H*	l, monoclinic	l	l	tilted to side b	no helix

s = short-range order; l = long-range order.

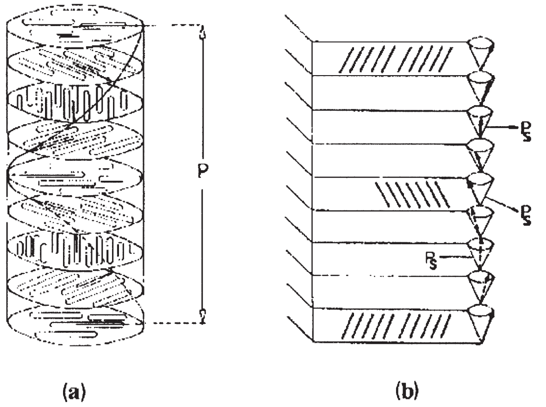


Figure 3
Helically twisted structures: (a) chiral nematic phase, (b) chiral smectic C* phase.

phases. By mechanical twisting of nematic layers made from non-chiral compounds, stabilized by supporting substrates, the layers have properties similar to low-twisted chiral nematics. They are the basis of the widely used electrooptic displays with twisted structure (TN-displays, STN-displays).

5.2 Smectic Phases of Chiral Compounds

The structures of the chiral smectic A and B phases are equal to those of the non-chiral variants, but due to the chirality of the molecules the phases possess additional features like very weak optical activity and the electroclinic effect.

Smectic C phases derived from chiral compounds possess unique properties. In addition to the tilted C phase structure, in adjacent layers the projection of the tilt of the layer plane is turned by a small angle, thus producing a twisted structure (Fig. 3(b)). This twist causes properties similar to those of twisted nematics, that is, strong optical activity and selective reflection of circularly polarized light, and special textures are also formed. When the compounds possess a lateral electric dipole component (which is the normal case), spontaneous polarization (P_S in Fig. 3(b)) develops within the surface layer. Because the orientation of this spontaneous polarization follows the twist of the structure, the spontaneous polarization is cancelled (helielectricity). By external fields or by interaction with supporting walls, the twist of the structure can be unwound and the structure shows ferroelectric properties. The direction of the macroscopic spontaneous polarization depends on the sense of the chirality of the molecules. These ferroelectric properties are the basis for fast switching displays (surface-stabilized ferroelectric liquid crystal (SSFLC) displays). It was in Sm C* phases that ferroelectric properties of liquid crystals were first predicted and detected.

There are also twisted variants of smectic I, F, and M phases occurring in chiral compounds Sm I*, F*, and M*. In the untwisted form these phases are ferroelectric.

In some chiral compounds the adjacent layers of Sm C* phases are tilted in the opposite direction, forming Sm C_A^* phases comparable to Sm C_{alt} phases. In addition the tilted layer pairs are in a twisted arrangement. The Sm C_A^* phases are antiferroelectric and using electric fields they can be switched to ferroelectric C* phases. They can be used for the construction of antiferroelectric liquid crystal displays (AFLCDs). There are also phases, in which the layers alternate between ferroelectric and antiferroelectric arrangement, which produce Sm C*(ferri) phases with ferrielectric character (often designated Sm $C\gamma^*$ phases). Because different mutual arrangements of the layers are possible, different types of ferrielectric phases exist.

Antiferroelectric properties have been found also in Sm I_A^* and Sm F_A^* phases, which possess alternating directions of tilt in adjacent layers.

5.3 Soft Crystal Phases of Chiral Compounds

In chiral compounds all known phase types of soft crystals can exist. Not all these phases possess the helical structure. The tilted phases J*, G*, K*, and H* are ferroelectric. Because it is very difficult to achieve homogeneous orientation of these phases, the investigation of their structures is still incomplete.

6. Liquid Crystals with Cubic Structure

6.1 Blue Phases

In a small temperature region between the chiral nematic and isotropic liquid phases in chiral calamitic compounds up to three different blue phases (BP I–III) can appear. They are called blue phases because of scattering of visible light in the blue or green region.

The blue phases have the following symmetries: BP I body centered cubic; BP II simple cubic; BP III amorphous. The blue phases only occur in compounds with strong chirality, and their structure is based on defect systems containing twist in two directions (Collings and Patel 1997). From the optical standpoint, they do not show birefringence. In electric fields several additional blue phases have been observed.

6.2 Cubic Phases

There are a considerable number of calamitic compounds that exhibit different cubic phases with liquid crystalline character (Diele and Göring 1998). These phases possess liquid-like molecular packing with high mobility of the molecules, and a three-dimensional periodicity of the lattice units consisting of associations of molecules.

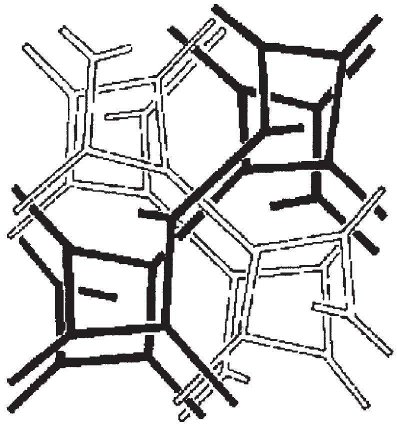


Figure 4
Model for the cubic structure of the space group *Ia*3*d*.

The commonly occurring space groups of these phases are *Ia*3*d*, *Pn*3*m*, and *Im*3*m*. It is not easy to find a space filling model for these cubic structures, which consist of rod-like elongated molecules. In most cases the cubic lattice parameter is larger than the molecule length, and the unit cells contain many molecules, in some cases up to several hundred. For the simplest form of the lattice units, bundles of molecules with effectively spherical shape are assumed. On the other hand, for cubic lyotropic liquid crystals several models with complicated three-dimensional interwoven networks of cylinders have been developed, which have been applied to thermotropic cubic phases as well. Figure 4 shows the model for the space group *Ia*3*d*, containing two interwoven labyrinths, each of which consists of rods linked three-by-three. It is still a matter for discussion how this network can be formed by the elongated molecules in a reasonable and space-filling manner.

7. *Rule of the Phase Sequences and Reentrant Behavior*

For the liquid crystalline and soft crystal phases a rule of the sequence of the different phase types can be derived from the observation of a large number of examples (Sackmann and Demus 1966, 1973). Starting from the solid crystalline state, a stepwise decrease of the order during heating is connected with the occurrence of the phases in a regular sequence

crystal| H, K, E, G, J| F, B, (I, B_{hex})|
crystal| *soft crystal*| *Sm hexatic*|

|N| isotropic| C_{alt}, C A|
|*N*| *isotropic*| *Sm*|

In the second line the groups of phases are indicated, which clearly show the stepwise decrease of order according to the rule of the phase sequences. It should be emphasized that the reverse sequence appears with decreasing temperature.

Substances can have a high degree of polymorphism: up to six liquid crystalline and soft crystal phases have been found in the same compound. But there is no compound with the complete general phase sequence. Despite gaps with respect to the general phase sequence, the mutual sequence of the phases generally is maintained.

The cubic phases do not fit in this rule, in most cases they appear below or above Sm C as adjacent phases.

In exceptional cases the phase sequences can be partly reversed, and certain phase types may exist in two or more separate temperature regions divided by regions of another phase type. This phenomenon mainly occurs with nematic and Sm A phases and is called "reentrant" (Cladis 1998). In simple reentrant cases the phase sequence solid–N_{re}–Sm A–N–isotropic has been found. But there are also cases of double or even triple reentrance, e.g., solid–Sm A_{re}–N_{re}–Sm A_{re}–N_{re}–Sm A–N–isotropic.

Reentrance relatively often occurs in compounds with strongly polar groups and in these cases it is explained by the temperature-dependent effective molecule shape of the constituent compounds. In other cases temperature-dependent steric effects seem to cause reentrance. In chiral compounds the rule of the phase sequence is also valid, and there are reentrant cases here as well.

8. *Concluding Remarks*

The system of phase types of liquid crystals is not complete or theoretically exact. The research in this area is still very active, and new phase types and subtypes of known phase types are expected to be found.

More recently, compounds have been synthesized that are on the boundary between rod-like and discotic shape. Some of these compounds can exhibit typical phases of calamitic compounds as well as typical columnar phases of discotic compounds, and in intermediate regions they also show cubic phases (Nguyen *et al.* 1998).

Another new field of research is banana-shaped (bent) molecules, which exhibit smectic-like phases different from those of ordinary calamitic molecules. Seven types of banana-phases are known (Pelzl *et al.* 1999).

There are molecules with board-like (sanidic) molecular shape. In such compounds biaxial nematic and biaxial Sm A phases have been claimed.

Other compounds with molecules deviating from the classical rod-like shape can also exhibit liquid crystalline phases of different structures (Demus 1998).

Bibliography

Chandrasekhar S 1998 Discotic liquid crystals. In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals*. Wiley-VCH, Weinheim, Germany, Vol. 2B, pp. 749–80

Cladis P E 1998 Re-entrant phase transitions in liquid crystals. In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals*. Wiley-VCH, Weinheim, Germany, Vol. 2B, pp. 391–405

Collings P J, Patel J S 1997 *Handbook of Liquid Crystal Research*. Oxford University Press, New York, pp. 99–124

Demus D 1998 Chemical structure and mesogenic properties. In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals*. Wiley-VCH, Weinheim, Germany, Vol. 1, pp. 133–87

Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) 1998 *Handbook of Liquid Crystals*. Wiley-VCH, Weinheim, Germany, Vols. 1–3

Demus D, Richter L 1978 *Textures of Liquid Crystals*, 2nd edn. VEB Deutscher Verlag für Grundstoffindustrie, Leipzig, Germany

Diele S, Göring P 1998 Thermotropic cubic phases. In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals*. Wiley-VCH, Weinheim, Germany, Vol. 2B, pp. 887–900

Goodby J W 1998 Phase structures of calamitic liquid crystals. In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals*. Wiley-VCH, Weinheim, Germany, Vol. 2A, pp. 3–22

Gray G W, Goodby J W 1984 *Smectic Liquid Crystals, Textures and Structures*. Leonard Hill, Philadelphia

Nguyen H-T, Destrade C, Malthete J 1998 Phasmids and polycatenar mesogens. In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals*. Wiley-VCH, Weinheim, Germany, Vol. 2B, pp. 865–85

Pelzl G, Diele S, Weissflog W 1999 Banana-shaped compounds—a new field of liquid crystals. *Adv. Mater* **11**, 707–24

Sackmann H, Demus D 1966 The polymorphism of liquid crystals. *Mol. Cryst*. **2**, 81–102

Sackmann H, Demus D 1973 The problems of polymorphism in liquid crystals. *Mol. Cryst. Liq*. **21**, 239–73

Seddon J M 1998 Structural studies of liquid crystals by x-ray diffraction. In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals*. Wiley-VCH, Weinheim, Germany, Vol. 1, pp. 635–79

D. Demus
Halle, Germany

Carbon Mesophase

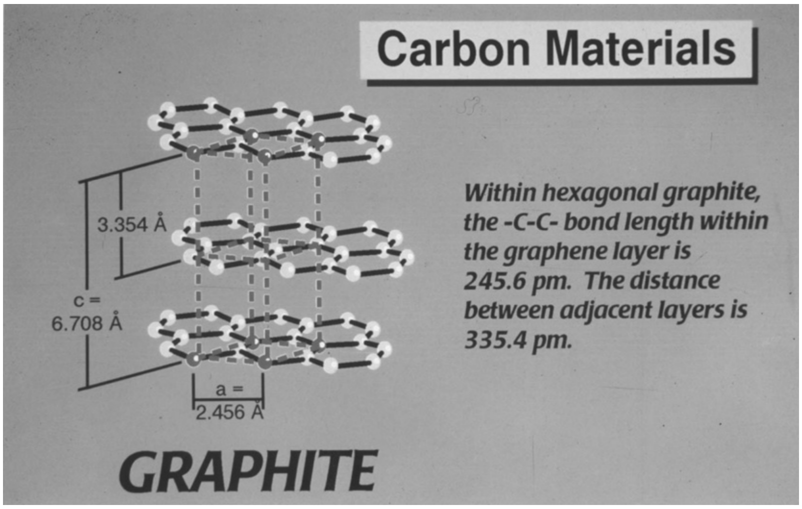
Carbon materials have important roles both domestically and industrially (Marsh *et al.* 1997). The common pencil has a graphitic point. The daily newspaper is printed with a carbon black ink. The common torch battery has a centrally placed carbon electrode. The lithium cell of the mobile phone operates with lithium intercalated into carbon electrodes (Chang *et al.* 1999). Sports equipment is constructed around carbon/carbon fiber composites and a modern passenger jet has structural components made of carbon/carbon fiber systems. High-purity graphite electrodes are used in the steel industry. Carbon anodes, made from delayed cokes from petroleum, are central to the aluminum industry, for the Hall–Héroult electrolysis cell. Part of nuclear energy generation is based on neutron graphite-moderated systems. Microporous carbons, with their immense adsorption capacities, have important purification roles. Metallurgical coke, from coal, is central to iron and steel production. The application of a particular carbon is determined, absolutely, by its structure. Carbon materials fall into two categories, the graphitizable, anisotropic carbons and the nongraphitizable, isotropic carbons. This article concentrates on the creation of structure in the graphitizable, anisotropic carbons.

1. *Structure in Carbons*

Single-crystal graphite has a structure of sheets or graphene layers of carbon atoms arranged in hexagons, the graphene layers being stacked below and above each other, as shown in Fig. 1. Defect-free, single-crystal graphite does not exist, industrial graphitic materials being defective derivatives of the perfect crystal. That is, the graphene layers possess defective hexagonal arrangements with, for example, atomic vacancies within the layer. In graphitizable, anisotropic carbons, i.e., those carbons which exhibit three-dimensional crystallinity based on the graphite lattice on heat treatment to above 2000 °C, the graphene layers (sized $>10\,\text{nm}$) are approximately parallel to each other. In nongraphitizable, isotropic carbons the graphene layers are much smaller (1.0 nm), and they are also quite defective, the graphene layers being arranged relative to each other in a random way. This is the structure of the isotropic microporous carbons. So how is it that carbons can exhibit such a wide range of structure? The answer lies in the carbonization procedures.

2. *Carbonizations*

It is an everyday experience that almost all naturally occurring organic materials, such as wood, nut shells, corn husks, and low- and high-rank coals, are transformed on heating in an inert atmosphere to about 800 °C into chars, charcoals, or carbons. No melting occurs in such systems and the resulting carbon structures are distorted copies of the structure of the parent organic material. Such carbons are the isotropic carbons. For the graphitizable, anisotropic carbons, the situation is completely different. The starting materials are petroleum pitch, coal-tar pitch, coal, and polynuclear aromatic chemicals such as naphthalene. During carbonization these starting



Carbon Materials

Within hexagonal graphite, the -C-C- bond length within the graphene layer is 245.6 pm. The distance between adjacent layers is 335.4 pm.

GRAPHITE

Figure 1
Diagram of the hexagonal unit cell of graphite.

materials pass through a fluid/liquid phase, the viscosity decreasing rapidly with increasing temperature (as is normal for liquids) until, quite suddenly, the system solidifies to a coke, anisotropic in structure, with a parallelism of graphene layers ensuring graphitization on further heat treatment above 2000 °C. What is happening here? The understanding of these phenomena was a major breakthrough for carbon science, and the story goes back to 1964 in Australia (Brooks and Taylor 1968). These authors introduced the concept of mesophase to carbon science and with it an understanding of properties of ‘‘liquid crystal systems.’’ The carbonization chemistry of petroleum pitch, etc., produced growth of constituent polynuclear hydrocarbons to sizes of >2000 amu, which then ‘‘self-assembled’’ to create the anisotropy of the solid coke product. Thus, the carbonization system was one of synthesis of mesogens (liquid-crystal-forming molecules) which, once formed, imparted the properties of bulk liquid crystals to the carbonizing system.

3. Liquid Crystals

Liquid crystals have been recognized since 1888 (see *Liquid Crystal: Overview*). They possess much more order than is found in normal Newtonian liquids (Demus 1994, Khoo 1995). This is because mesogens possess sufficient polarity for cohesion to be maintained between molecules when in the fluid phase. Polarity can result from nitrogen and oxygen groups in synthesized rod-like (rodic) molecules, or from the π–π interactions between disk-shaped (discotic) polynuclear aromatic mesogens. As for a solid, with increasing temperature a liquid-crystal compound has a well-defined transition temperature from a solid to a liquid crystal. The cohesion between molecules is sufficiently strong to resist the forces of separation associated with the kinetic energies of the molecules. With increasing temperature the viscosity decreases until a second transition occurs from the liquid-crystal phase to an isotropic Newtonian liquid. This is the behavior of *nematic* liquid crystals where mesogens are organized within parallel planes, but with no organization above and below the planes. On the other hand, there exist *smectic* liquid crystals in which there is order between molecules both within planes and above and below the planes. As for a solid, with increasing temperature the smectic liquid crystals exhibit three transition temperatures, from solid to a smectic phase, from smectic to a nematic phase, and from the nematic phase to the isotropic Newtonian liquid phase. The liquid-crystal systems from pitch carbonizations are of the discotic, aromatic, nematic type, with transient smectic behavior (Fig. 2) (Greinke 1994, Lewis and Singer 1980, Oberlin *et al.* 1999).

4. Liquid Crystals from Pitch Materials

There are three principal differences between carbonaceous discotic, aromatic, nematic liquid crystals (mesophase) and conventional rodic nematics (Marsh and Diez 1993, Marsh *et al.* 1999).

(i) Conventional liquid crystals are chemically stable whereas carbonaceous discotic, aromatic nematic liquid crystals (mesophase) are chemically reactive. Once formed, the polymerization processes of molecular growth, within carbonizing pitch, continue within planes and between planes.

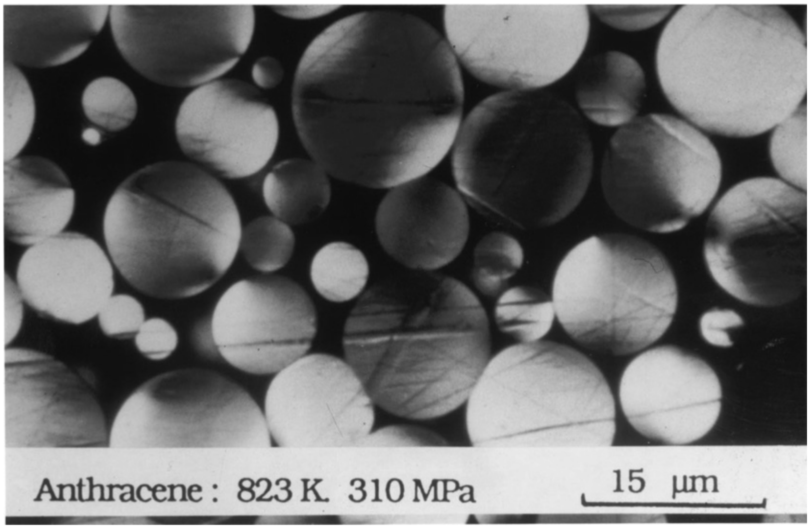


Figure 2
Optical micrograph of mesophase spheres from anthracene, 550 °C, 310 MPa.

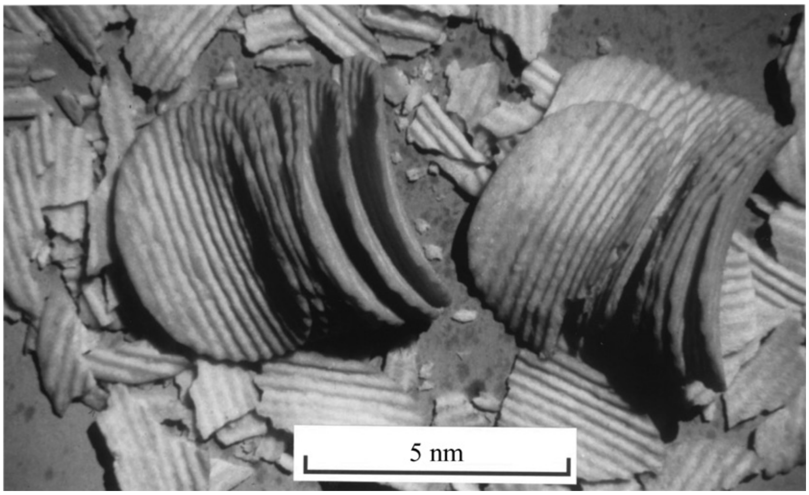


Figure 3
Model of self-assembly of mesogens in carbonizing pitch.

(ii) If the temperature of newly formed mesophase is very rapidly increased, over a 50 °C interval, then the liquid crystal/isotropic liquid transition is observed. On cooling, the liquid crystal reappears. With a subsequent re-heat, again the liquid crystal/isotropic liquid transition is observed, but at a higher temperature, as a result of continuous increases in the molecular weight of constituent mesogens. That is, there are no constant transition temperatures associated with mesophase.

(iii) During carbonizations, with increasing temperature mesophase formation is always running ahead of the mesophase/isotropic phase transition temperature. As a result, irrespective of heating rates, the end product is an anisotropic carbon and not a disorganized carbon.

Care needs to be taken when describing mesophase growth processes. It is not a process of crystallization, nor is it a precipitation process. Correctly, it is a process of molecular self-assembly, an important property of macromolecules (see Fig. 3) (Marsh *et al.* 1999).

5. *Observations of Mesophase*

The first observations of mesophase were made by Brooks and Taylor (1968) in a coal seam which had

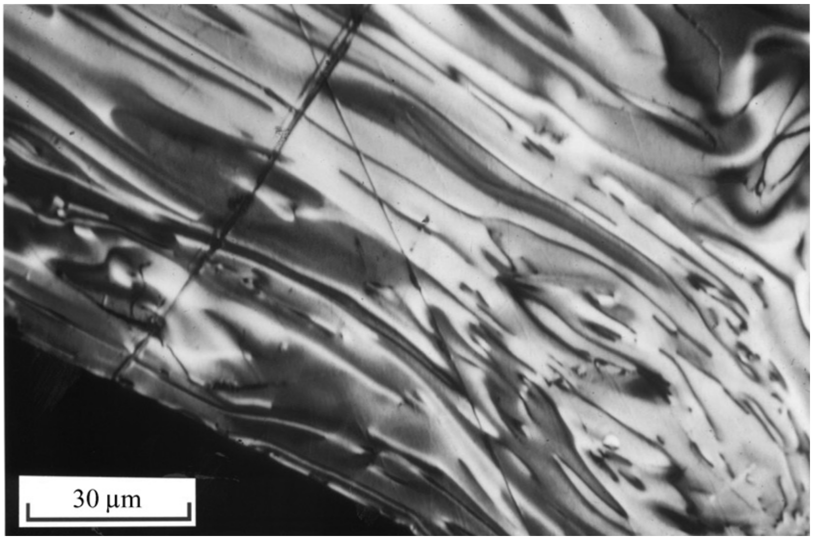


Figure 4
Optical micrograph of coalesced mesophase from a petroleum pitch.

been heated over a period of months by a dyke of magma at about 1000 °C. Direct observation of polished surfaces of the coal samples by polarized light microscopy revealed reflecting spheres, 1 μm in diameter, within the coal matrix. Within the laboratory, using polarized light optical microscopy equipped with a hot-stage, a carbonization of a petroleum pitch could be observed directly at heat treatment temperatures $>375\,^{\circ}\mathrm{C}$. Initially, anisotropic spheres, $\sim 1.0\,\mu\mathrm{m}$ diameter, were observed (Fig. 2). These grew in size to about 10 μm diameter when a process of coalescence rapidly increased the sizes of the anisotropic growth units until all of the sample became anisotropic (bulk mesophase) (Fig. 4). It soon became apparent that not all pitch samples—various petroleum residues, coal-tar pitches, and model aromatic compounds—behaved similarly. For some, coalescence was complete, producing a continuous single mesophase system. For others, the spheres would grow, but on contact there was no coalescence; they simply adhered together to produce a multicrystalline mesophase. The implications of these observations for the carbon industry were considerable. Graphitization of a single mesophase system would produce a better graphite than a multicrystalline mesophase. It was also appreciated that a mesophase behaves viscoelastically, that the shapes of single- and multicrystalline mesophases could be changed by physical deformation. Thus, the science of mesophase growth, size, and shape opened up methods of producing graphites with specialized applications, such as "isotropic" graphites for the nuclear industry.

6. *Structure of Mesophase*

Structure within the growth spheres of mesophase controls the important property of coalescence. X-ray diffraction studies of the spheres indicated that the discotic mesogens were stacked parallel to each other, the stacking being parallel to an equatorial plane. As result of this, the edges of the mesogens are located on the surface of the spheres. An alternative is where the mesogens wrap themselves around the surface of the sphere in an onion-like way. The phenomenon of equatorial stacking facilitates coalescence of spheres on physical contact, the mesogens of each sphere sliding into each other to create a larger sphere but with the same equatorial stacking. Spheres with onion-like assemblies of mesogens could possibly not have coalesced so easily. The entire process of the formation of graphitic carbons thus depends upon coalescence of mesophase growth units.

The optical microscope has a resolution of about 1 μm, and this factor raises the question of whether or not mesophase exists before it can be seen by optical microscopy. Phase-contrast, high-resolution electron microscopy helped to answer this problem (Oberlin 1989, Oberlin *et al.* 1999). This technique is powerful enough to resolve individual mesogen molecules as fringes (Fig. 5). From lattice-fringe micrographs the lengths of the fringes (relatable to mesogen size) as well as the number of fringes in a stack can be measured. Using a laser-beam optical-bench the relative misorientation of fringes and the interfringe spacings can be measured.

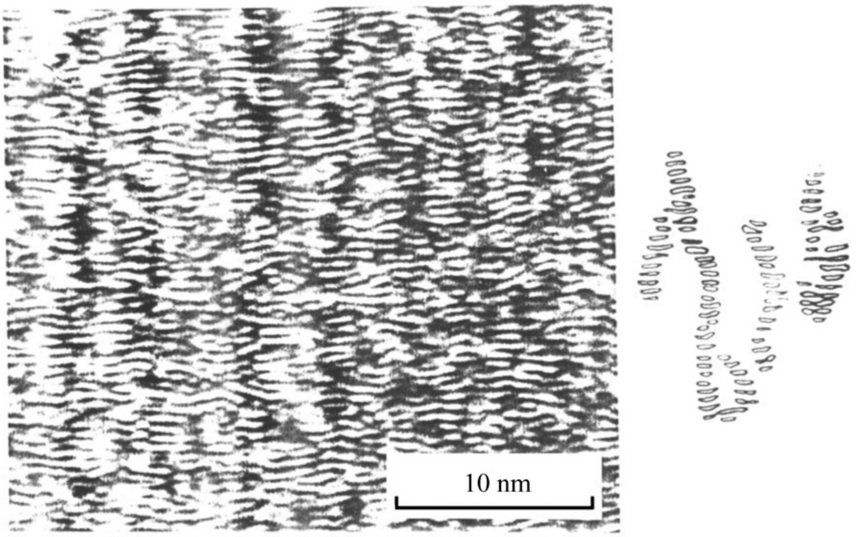


Figure 5
Phase-contrast, high-resolution, transmission electron micrograph of a thin section of a single noncoalesced mesophase sphere, showing fringe-images relating to mesogen locations, and inset, columnar stacking of mesogens as found in smectic liquid crystals (after Oberlin 1989).

Three results of importance are derived from these studies. First, the mesogens could be resolved as fringes at a stage in the carbonization system when no anisotropy was visible by optical microscopy. Second, columnar stacking of mesogens existed in the single noncoalesced sphere, this arrangement being reminiscent of smectic properties. This columnar stacking degenerated on coalescence of spheres to that of a nematic system. Third, as the mesophase crystallinity (seen in optical microscopy as optical texture) decreased in size, so the fringes became more distorted (less planar) resulting in larger interplanar fringe (mesogen) distances caused by out-of-plane defects. These combined microscopy techniques confirm totally the mechanisms of conversion of pitch-like substances to the liquid-crystal mesophase to the graphitizable carbon.

7. The Chemistry of Mesophase Formation

The viscosity/temperature curves for the pyrolysis of feedstocks, e.g., petroleum pitches, coal-tar pitches, polyvinyl chloride, and model aromatic compounds, indicate significant differences (Marsh *et al.* 1999, Marsh and Walker 1978). The most ordered of mesophases, showing optimum coalescence and providing the most graphitizable of carbons, had the lowest viscosities (highest fluidities) over the widest ranges of carbonization temperatures. The opposite effects were true—high minimum viscosities, over short temperature intervals, provided the least graphitizable carbons. Clearly, the graphitizability of a resultant carbon, this being an indication of the commercial value of the feedstock, is a function of the viscosity of the system. Viscosity is a function of molecular size and molecular cohesion, and these in turn are controlled by the chemistry occurring within a system, and are a function of chemical composition. Thus, there is established, via mesophase theory, a direct correlation between chemical composition of a feedstock and the graphitizability of the resultant carbon/coke.

Electron spin resonance studies indicate the presence of unpaired electrons in carbonization systems, and involve the formation of free radicals. The higher the "reactivity" of these radicals, the earlier will molecular growth occur. However, early molecular growth is associated with lower temperatures and higher viscosities, resulting in an inability of the highly viscous mesophase spheres to coalesce, thus producing carbons with a lower ability to graphitize. The presence of hydrogen-donor molecules promotes graphitizability. The hydrogen-donor molecule, such as a dihydronaphthalene, transfers a hydrogen atom to a reactive free radical, stabilizes the system to enable higher temperatures and lower viscosities to be reached, thus promoting mesophase coalescence and graphitizability of resultant carbons.

Heteroatoms, such as sulfur, oxygen, and nitrogen, and functional groups of molecules in the carbonization system enhance chemical reactivity and consequently lower the abilities of resultant carbons to graphitize. Metals catalyze carbonization reactions and so influence the development of optical texture (a measure of macrocrystallinity as seen in the optical microscope). The alkali metals inhibit the development of mesophase by enhancing reactivity too much. On the other hand, systems that are so unreactive that no formation of mesophase is possible, can be catalyzed to produce mesophase. Some effective catalysts include $AlCl_3$, ferrocene, and HF/BF_3.

8. *Industrial Relevance of Mesophase*

The awareness that structure in the many forms of commercial graphitizable carbons is a consequence of different development modes of the mesophase made a tremendous impact on the carbon/graphite-producing industries. Not only was the scientific control of established procedures improved, but new processes were also created.

8.1 The Delayed Coker

The piece of equipment used in industry to convert feedstocks to a graphitizable carbon (green coke) is the delayed coker (Mochida *et al.* 1994). This consists basically of a drum which is charged with several hundred tonnes of preheated petroleum feedstock ($>300\,°C$), subsequently heated further to 450–475 °C, and kept at this temperature for about 24 hours to produce the green coke. The quality of this green coke, in terms of optical texture, is of critical importance. The larger the size of the optical texture, the higher its commercial value. There exist two qualities of commercial green coke: *regular* or *sponge* coke with optical textures of 10 μm in diameter (coarse-grained mosaics) and medium-flow anisotropy ($<30\,\mu m$ in length); and *needle* coke with acicular flow domain anisotropy ($>60\,\mu m$ in length). The calcined regular cokes are the filler cokes in anode production for use in the Hall–Héroult cell to produce aluminum from the electrolytic reduction of aluminum oxide. Calcined needle cokes act as filler cokes in the manufacture of high-performance graphite electrodes used to transmit electrical power to crucibles for steel production.

8.2 Metallurgical Coke

Metallurgical coke is used in blast furnaces. Its role there is threefold: to support the iron ore burden (high strength); to act as the carbon source for the reduction of iron oxide ores (suitably low reactivity); and to provide thermal energy. Metallurgical coke must have an optical texture of fine-grained mosaics (1.5 μm in diameter). The theory of carbonization of coals, including the use of coal blends and of additives (hydrogen donors), has benefited the coal-coking industry considerably.

8.3 Carbon Fibers

The concept of spinning mesophase into carbon fibers, to take advantage of the prealignment of molecules in the liquid-crystal structure to obtain high-strength, high-modulus fibers, has been actively pursued, mainly in Japan. Initially, mixtures of pitch with mesophase were spun, but spinning conditions were extremely difficult to control (Greinke 1994, Blanco *et al.* 1999). A pitch-like material, a mesophase pitch, completely anisotropic by optical microscopy, has been synthesized from the pure aromatic hydrocarbons naphthalene and methylnaphthalene with the aid of HF/BF_3. These synthesized mesophase pitches have many advantages, such as low softening point and viscosity, when compared with such conventional pitches as petroleum pitches and coal-tar pitch, prepared by pyrolysis processes. Mitsubishi Gas Chemical Co. constructed a semicommercial facility (1000 tonnes per year) at their Mizushima plant using HF/BF_3. The spinning of mesophase pitch has many advantages over conventional pitches. The shape and size of the spinneret nozzles and spinning rate all influence structure within the fiber (Hong *et al.* 1999). Complex morphologies resulting from such spinning, perhaps unexpected but beneficial, can be seen in SEM micrographs of fracture surfaces of such fibers. A new dimension has been added to carbon fiber technology.

8.4 Carbon Blacks

These soot-like particles are formed by the vapor-phase pyrolysis/partial combustion of hydrocarbons (Marsh *et al.* 1997). The mechanism of formation involves the generation of liquid droplets (clusters of molecules). For some carbon blacks, the establishment of the concentric, onion-like arrangement of constituent molecules within the droplet, with parallelism of the structure to the surface of the droplet, occurs most probably via a liquid-crystal mechanism.

Bibliography

Blanco C, Fleurot O, Menéndez R, Santamaria R, Bermejo J, Edie D D 1999 Contribution of the isotropic phase to the rheology of partially anisotropic coal-tar pitches. *Carbon* **37**, 1059–64

Brooks J D, Taylor G H 1968 In: Walker Jr P L (ed.) *Chemistry and Physics of Carbon*. Dekker, New York, Vol. 4

Chang Y-C, Sohn H-J, Ku C-U, Wang Y-G, Korai Y, Mochida I 1999 Anodic performances of mesocarbon microbeads (MCM8) prepared from synthetic naphthalene isotropic pitch. *Carbon* **37**, 1285–97

Greinke R A 1994 In: Thrower P A (ed.) *Chemistry and Physics of Carbon*. Marcel Dekker, New York, Vol. 24

Hong S-H, Korai Y, Mochida I 1999 Development of mesoscopic textures in transverse cross-sections of mesophase pitch-based carbon fibers. *Carbon* **37**, 917–30

Lewis I C, Singer L S 1980 In: Walker Jr P L, Thrower P A (eds.) *Chemistry and Physics of Carbon*. Dekker, New York, Vol. 17

Marsh H, Diez M A 1993 In: Shibaev V P, Lam L (eds.) *Liquid Crystalline and Mesomorphous Polymers*. Springer, New York

Marsh H, Heintz E A, Rodriguez-Reinoso F 1997 *Introduction to Carbon Science*. Universidad de Alicante, Secretariado de Publicaciones, Spain

Marsh H, Martinez-Escandell M, Rodriguez-Reinoso F 1999 Semicokes from pitch pyrolysis: mechanisms and kinetics. *Carbon* **37**, 363–90

Marsh H, Walker P L Jr 1978 In: Walker Jr P L (ed.) *Chemistry and Physics of Carbon*. Dekker, New York, Vol. 15

Mochida I, Fujimoto K, Oyama T 1994 In: Thrower P A (ed.) *Chemistry and Physics of Carbon*. Dekker, New York, Vol. 24

Oberlin A 1989 In: Thrower P A (ed.) *Chemistry and Physics of Carbon*. Dekker, New York, Vol. 22

Oberlin A, Bonnamy S, Rouxhet P G 1999 In: Thrower P A, Radovic L R (eds.) *Chemistry and Physics of Carbon*. Dekker, New York, Vol. 26

H. Marsh
North Shields, UK

Carbon Nanotubes

In the history of chemistry, the landmark events have doubtless been the discovery of various elements, and subsequently the various allotropes of some of these elements. The 1985 discovery of fullerenes (the third ordered allotrope of carbon after graphite and diamond) by Kroto *et al.* (1985) was in itself an historic event considering that it was the only allotrope of any element to have been discovered in the twentieth century. The level of excitement that these new materials generated in the chemistry, physics, and materials science communities was, needless to say, extremely significant. Another key milestone in fullerenes research was the production of experimental quantities of C_{60} and C_{70} by the evaporation of graphitic electrodes in a direct current (DC) arc-discharge environment by Krätschmer *et al.* (1990). The discovery of carbon nanotubes by Iijima (1991) was in turn intimately tied in with the use of the DC arc-discharge for fullerene synthesis. While the conventional fullerenes were formed in the vapor phase of the arc-discharge process and are mixed with the amorphous carbon deposited on the arc chamber walls, Iijima demonstrated that the core of the cylindrical growth on the face of the cathode contained some exotic nanostructures of carbon. The most striking among these structures were the carbon nanotubes—which were long hollow fibers with multiple layer planes on either side of the core, easily imaged by high-resolution transmission electron microscopy (TEM). Such an electron micrograph of a defect-free, multiwalled carbon nanotube is shown in Fig. 1.

1. *Multiwalled Carbon Nanotubes*

In reality, the existence of hollow multilayered filaments of carbon with diameters in the nanometer scale had been known for several decades prior to Iijima's discovery. Such filaments were formed by the catalytic action of nanometer-sized metal particles interacting with hydrocarbon gases inside reactors, and deposited in copious quantities on the reactor walls or tubes. Typically, such vapor-phase grown filaments were extremely defect-laden and they were viewed more as nuisance materials that affected the thermal and mass transport mechanisms inside the reactors and furnaces than as useful materials. On a comparative basis, the carbon nanotubes synthesized by Iijima by DC arc-discharge were finer and relatively free of imperfections. In fact, they can be considered analogous to miniaturized versions of the famous graphite whiskers discovered by Bacon in 1960, a schematic of which is shown in Fig. 2. It might be worthwhile to point out that the graphite whiskers synthesized by Bacon (whose diameters were typically in the 2–5 μm range) consisted of scrolls extending continuously along the length of the filament, the graphitic *c*-axis being exactly normal to the filament axis. Mechanical property measurements indicated that these whiskers possessed tensile strengths and modulus values of ~20 GPa and 1000 GPa, respectively.

In order to study these new materials in greater detail, the key first step was to be able to make them in macroscopic amounts. The manipulation of the conditions inside the arc chamber, such as current, electric field, and particularly the inert gas pressure was the key factor in achieving this step, as demonstrated by Ebbesen and Ajayan in 1992. However, the DC arc-discharge method of synthesizing carbon nanotubes also produced a plethora of side-products, such as multiwalled polyhedral particles, sheets of graphite, as well as small amounts of disordered carbon. A typical low-magnification, bright-field TEM image of the material obtained from the core of the cathode growth following an arc-discharge experiment is shown in Fig. 3—all of the different structures of carbon produced are illustrated in this image. If one desired to study the properties of pure carbon nanotubes, a purification step was essential. Among various methods used, one of the more successful routes to such a purification from the arc process was the preferential oxidation of the other forms of carbon to leave behind a sample that was composed only of nanotubes. The major drawback of this process was that more than 99% of the material was lost before one obtained a sample of pure nanotubes. A second approach, which results in higher yields of retained carbon nanotubes, involves the intercalation of the cathode product (which contains the nanotubes, nanopolyhedra, graphite, and disordered carbon) with materials such as copper chloride and their subsequent reduction and use of the resultant copper as

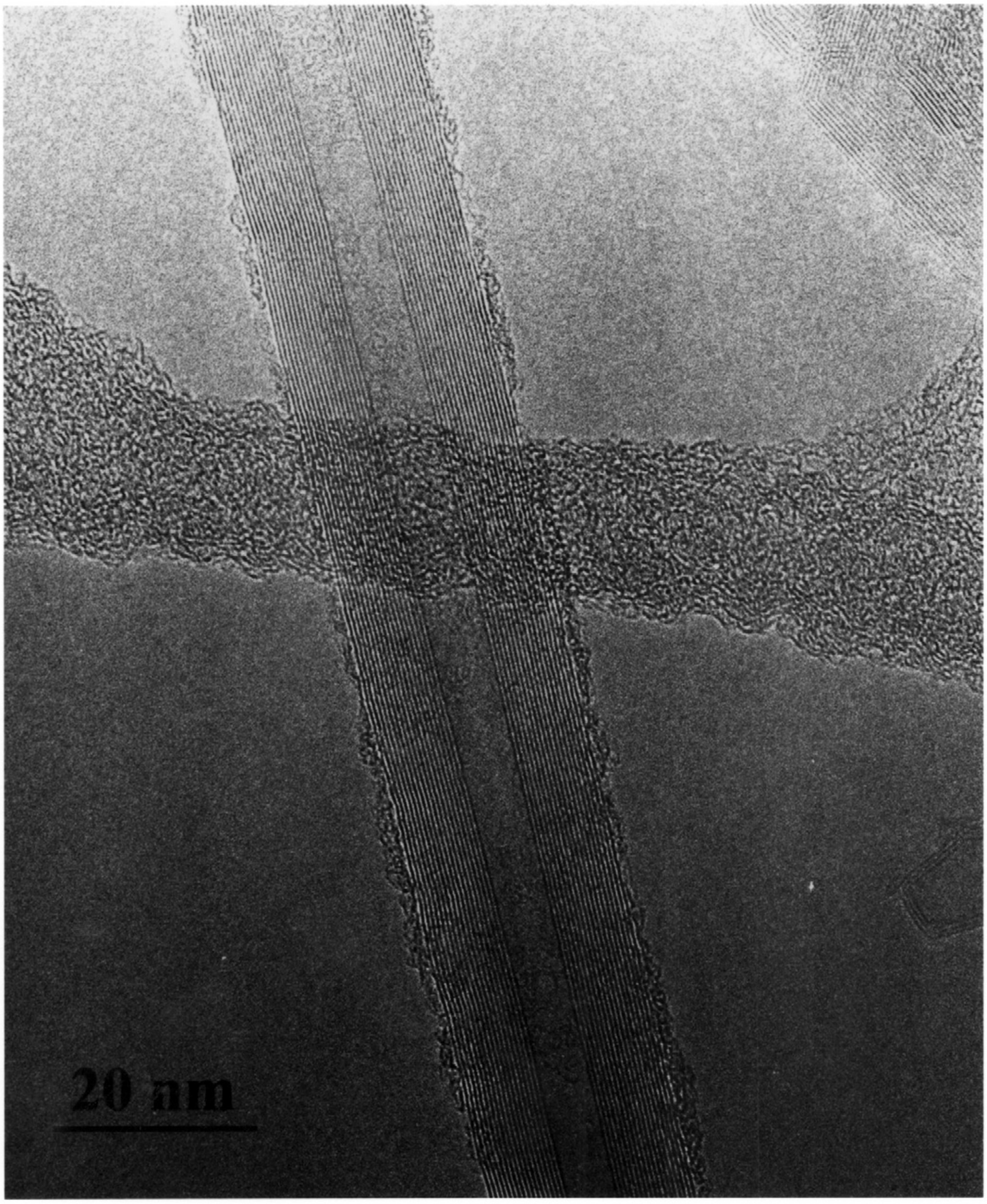


Figure 1
High-resolution TEM image of a relatively perfect region of a multiwalled carbon nanotube. The dark lines on either side of the core correspond to the basal plane of graphite.

an oxidation catalyst. Since the accompanying materials have a more open structure, they intercalate more easily and subsequent reduction of the chloride results in the preferential oxidation of the other forms of carbon. Other techniques such as the use of surfactants followed by filtration, chromatography, or centrifugation have also been attempted; it suffices to note that a technique where the process variables are difficult to control, such as arc-discharge, is nowhere near ideal for large-scale synthesis of nanotubes.

The microstructure of these newly discovered carbon nanotubes was clearly a very fascinating topic and characterizing them at the atomic level was of paramount importance to scientists the world over.

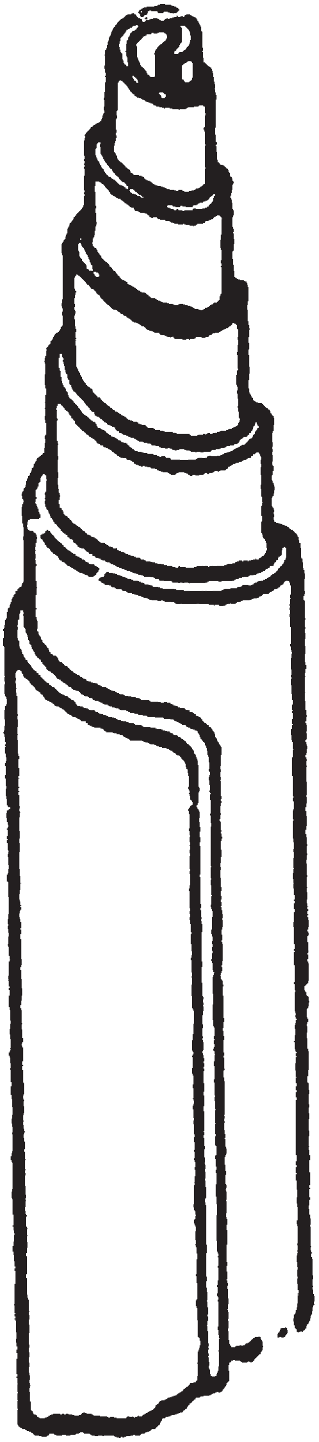


Figure 2
Schematic representation of a scroll-type graphite whisker (reproduced by permission of the American Institute of Physics from *J. Appl. Phys.*, 1960, **31**, 283–90).

Initially observed by high-resolution TEM, they were thought to be composed of seamless arrays of cylindrical sheets of graphene, which is defined as layer of graphite one atom thick and which corresponds to the basal 0002 plane of graphite, with aspect ratios generally >100 and with outer diameters mostly in the tens of nanometers. In other words, they could be essentially considered as single-dimensional objects with potentially exciting and hitherto unknown properties. Various analytical techniques have been employed by numerous investigators to understand how these tubes are constructed at the atomic or molecular level. Diffraction techniques using x-ray or electron sources, while not straightforward to interpret, have been invaluable in developing an understanding of the crystallographic details on a layer-to-layer basis, and scanning probe microscopy has also been helpful in studying the surface topography at the atomic level. However, high-resolution TEM has clearly been the technique of choice for several investigators to analyze various structural details such as tube stacking, defects, and internal and external closures associated with multiwalled carbon nanotubes.

Visual inspection of a large number of high-resolution TEM images of multiwalled carbon nanotubes clearly indicates that these materials have rather complex structures that are fraught with numerous defects. It is believed that the chaotic environment associated with the evaporation of graphite electrodes in an arc leads to these complex structures. An illustration of a ‘‘complex’’ structure is that of a multiwalled carbon nanotube with a uniform layer spacing on one wall and a variable layer spacing on the opposite wall (an eccentric rather than a concentric multiwalled nanotube), a high resolution TEM image of which is shown in Fig. 4. The most basic question regarding the structure of carbon nanotubes is whether the layer planes are all part of a single scroll (scroll type structure), or whether the multiwalled tube is composed of several discrete single-walled tubes nested one inside the next (nested structure). Structural analysis by several sets of investigators appears to indicate that most individual tubes consist of hybrid structures of scroll-like and nested features. These two types of structures are likely to co-exist either in the lateral sense or in the axial sense, and they would be typically separated from each other by crystallographic defects such as edge dislocations running parallel to the tube axis. Variable layer spacing, such as those illustrated for the ‘‘eccentric tube’’ in Fig. 4, are typically associated with nested tubes (such features would be difficult to reconcile with a scroll type structure). A high-resolution TEM image of a multiwalled carbon nanotube with numerous defects such as internal closures and pentagonal and heptagonal disclinations leading to asymmetric structures is illustrated in Fig. 5.

Based on the assumption that the in-plane Young’s modulus of graphite is 1000 GPa, it was natural for

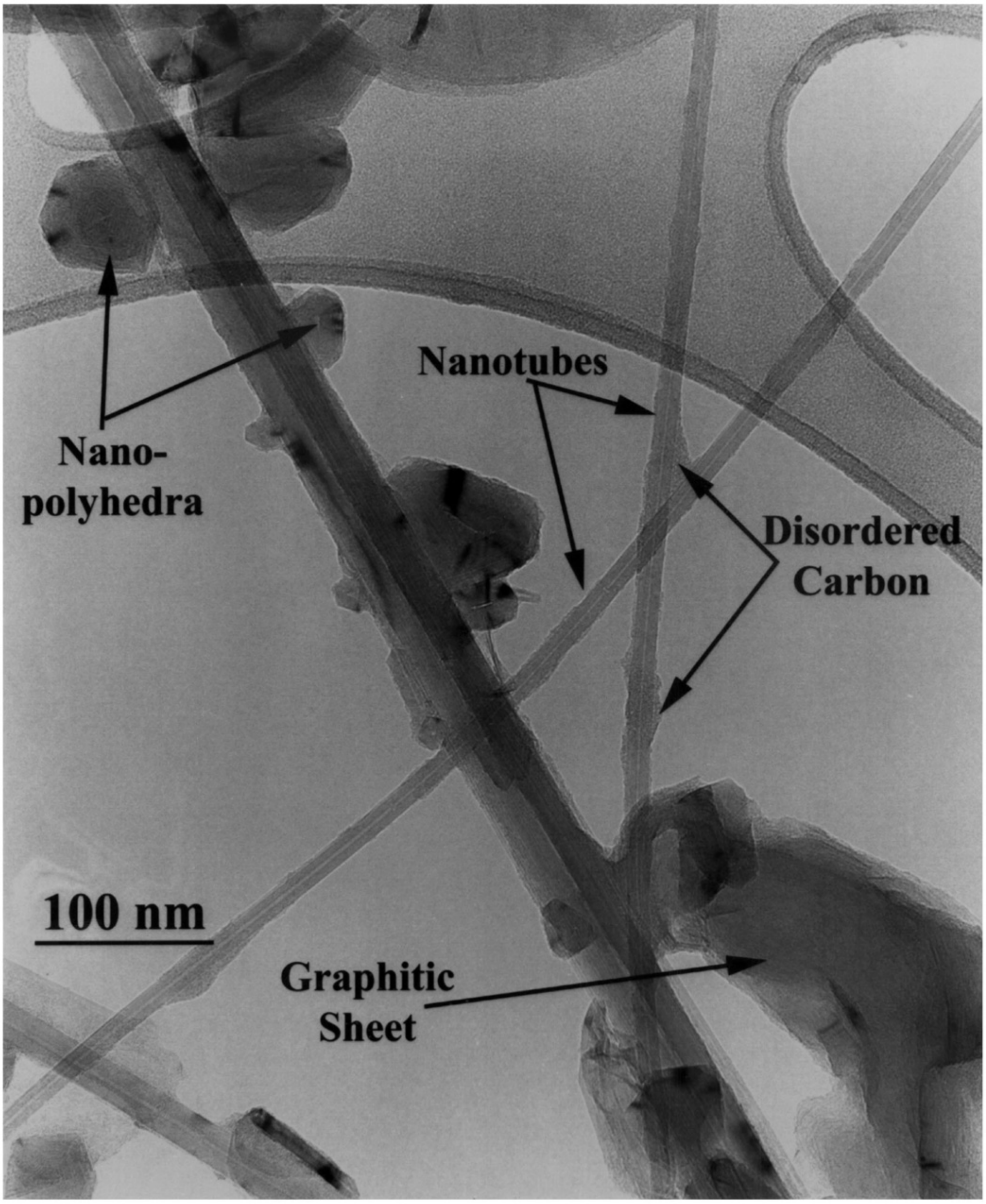


Figure 3
Bright-field TEM image of the material obtained from the cathode of a DC arc-discharge experiment showing carbon nanotubes, nanopolyhedra, graphitic sheets, and disordered carbon.

researchers to believe that the newly discovered carbon nanotubes would have extremely remarkable mechanical properties. Considering that they had a hollow core, the mechanical property-to-weight ratio of carbon nanotubes was also considered to be extremely impressive as well. While micrometer-sized carbon fibers are relatively easy to handle for the direct measurement of their mechanical properties using conventional laboratory equipment, the situation is quite the opposite with carbon nanotubes. Serendipity has played a significant role in offering some glimpses of the mechanical properties of carbon nanotubes when they were being examined by electron microscopy—their walls can be deformed

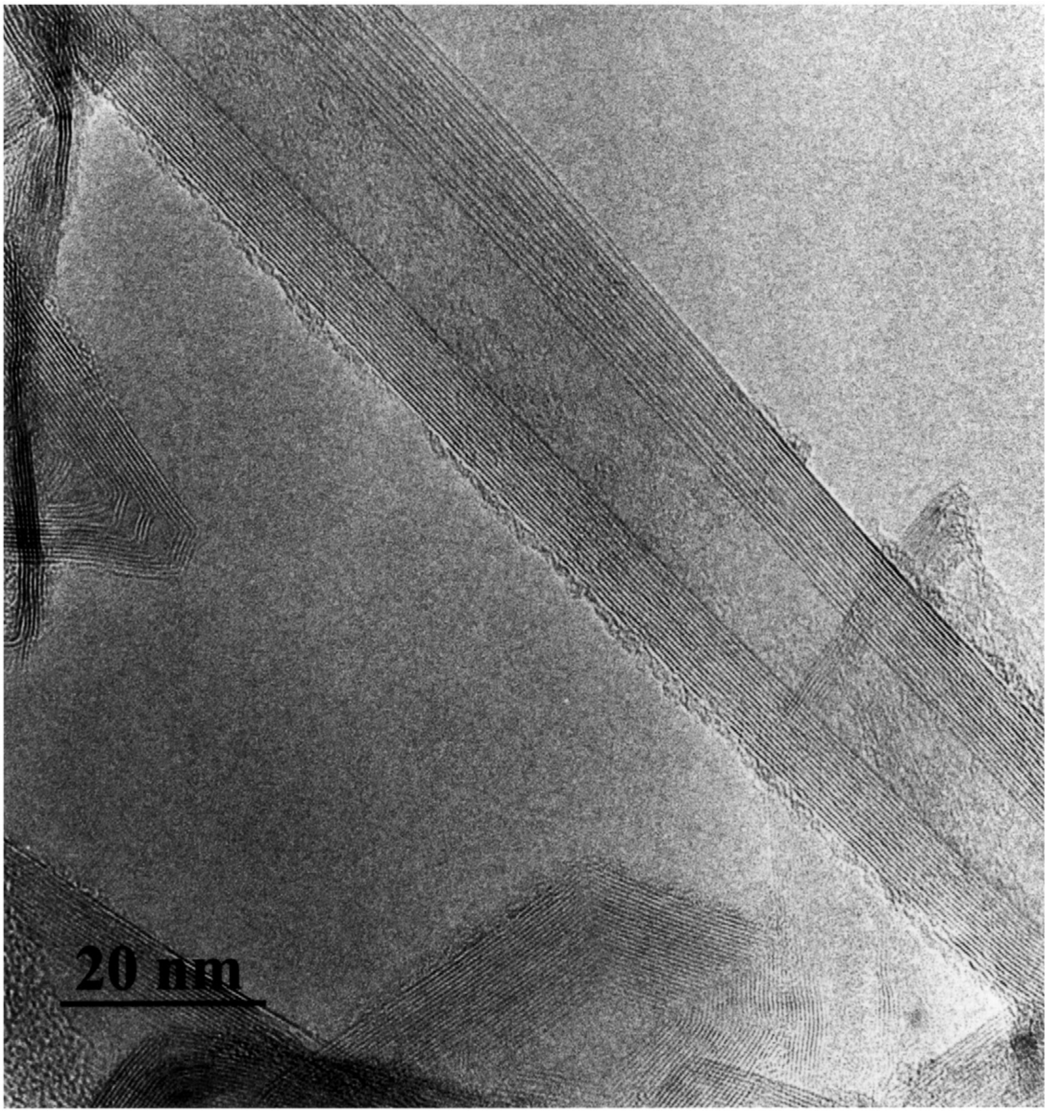


Figure 4
High-resolution TEM image of a multiwalled carbon nanotube showing variable interplanar spacing on one of its walls.

laterally by van der Waals forces generated by another tube lying side-by-side. On other occasions, tubes that were bent accidentally during sample preparation for TEM exhibited more damage on the compressive side than the tensile side. While direct measurement of the mechanical properties is by no means an easy and straightforward experiment, Yu *et al.* (2000b) have recently demonstrated that multiwalled carbon nanotubes could indeed be loaded in tension under *in situ* conditions inside a scanning electron microscope equipped with a custom designed tensile stage. While the degree of control in such experiments is far from ideal, their work has demonstrated that multiwalled carbon nanotubes typically fail by a "sword-in-sheath"-type fracture mechanism, similar to that observed in some carbon fibers. While the scatter in their data was rather significant, the observations of Yu *et al.* (2000a) suggest that the tensile strength of multiwalled tubes is typically on the order of several tens of gigapascals and that the Young's modulus is typically on the order of several hundreds of gigapascals.

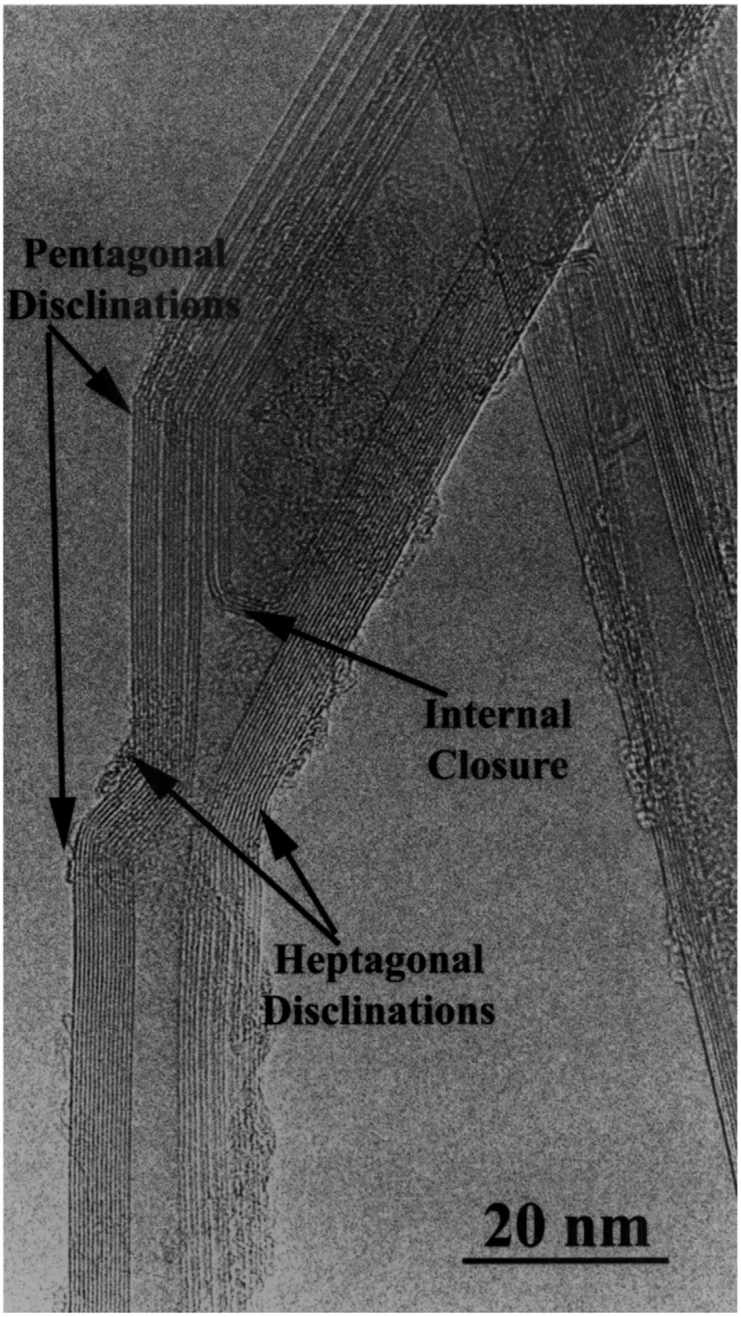

Pentagonal Disclinations
Internal Closure
Heptagonal Disclinations
20 nm

Figure 5
High-resolution TEM image of a multiwalled carbon nanotube showing features such as internal closures, pentagonal and heptagonal disclinations, and variable interplanar spacings.

Theoretical predictions regarding the electronic properties of carbon nanotubes were being made by several groups of researchers even prior to Iijima's spectacular discovery of multiwalled tubes in 1991. These predictions were typically made using elongated fullerenes as models, and hence they could be considered more analogous to single-walled tubes rather than multiwalled tubes. In these theoretical calculations, the helicity and diameter of the tubes were considered as the key factors controlling the electronic properties of nanotubes, such as whether they behaved as conductors or semiconductors. However, based on rather sophisticated lithographic techniques at the submicrometer level, Olk and Heremans (1994) have demonstrated that multiwalled carbon nanotubes behave similar to nanowires, with a density of states and an energy gap that are inversely proportional to the tube diameter. While these preliminary experiments appear to indicate that multiwalled tubes exhibit electrical resistance behavior similar to semiconductors, a major challenge involving the attachment of four nanoprobes to a single nanotube for more precise measurement of electronic properties has not been achieved to date.

2. *Single-walled Carbon Nanotubes*

As stated in the past section, several theoretical predictions were made regarding the unique electronic and mechanical properties of one-dimensional, tube-shaped fullerenes at or about the time of the discovery of multiwalled carbon nanotubes. In 1993, the excitement over the discovery of multiwalled nanotubes two years earlier was significantly eclipsed by the almost simultaneous discoveries at NEC in Japan (Iijima and Ichihashi, 1993) and at IBM in the USA (Bethune *et al.* 1993) of single-walled variants of the carbon nanotubes. Contrary to their multiwalled counterparts, these tubes were more similar to conventional fullerenes in the sense that they are formed in the vapor phase of the DC arc-discharge process. When transition metal catalysts were incorporated into the anode, the carbon that deposited on the chamber walls and other surfaces had a rubbery or web-like texture contrary to the conventional powdery texture one obtained when no catalysts were used in the anode. Examination of this odd-textured soot by TEM subsequently led to the discovery of the single-walled carbon nanotubes. The single-walled tubes typically grew as bundles (referred to by the research community as ropes), and individual tube diameters generally varied between 1.3 nm and 1.6 nm. A typical electron micrograph of single-walled carbon nanotubes is shown in Fig. 6.

The discovery of the single-walled carbon nanotube was extremely fascinating to the research community for several reasons. For one, they offered a possible route to experimentally verify several of the theoretical studies based on single-walled tubes as models. Secondly, the arc-discharge produced single-walled tubes that were extremely uniform in diameter compared to their multiwalled counterparts, and they were relatively free of defects as well. While the role of the transition metal catalyst in the formation of single-walled nanotubes was not immediately clear, follow-up research demonstrated that the content of single-walled nanotubes could be substantially enhanced in the soot by using mixtures of rather than single transition metals as catalysts. Finally, alternate methods such as the use of double-laser vaporization technique using transition metal mixtures as catalysts were also found to be capable of producing single-walled carbon nanotubes in abundant quantities.

Considering that single-walled carbon nanotubes were the quintessential single-dimensional objects, the structure and diameter of individual tubes are paramount in determining their physical properties. The structure of a single-walled carbon nanotube could be classified as either armchair, zigzag, or chiral, and these three structures are clearly depicted in the schematic diagram, reprinted from the work of Dresselhaus *et al.* (1995b), shown in Fig. 7. Subsequent analysis of the structure of these materials by techniques such as high-resolution TEM has shown that individual tubes are far from perfect, and, depending on the technique used in their processing, defect structures have been observed on tube walls as well as on their caps. Electron nano-diffraction studies on ropes of single-walled carbon nanotubes produced by arc-discharge or by laser ablation have also been carried out with the intent of determining the relative amounts of tubes of different chiralities. The overall assessment by various groups appears to indicate that individual ropes are typically composed of tubes ranging from zigzag to chiral to the (10, 10) armchair structure.

Theoretical analysis predicts that ideal single-walled carbon nanotubes would have extremely impressive mechanical properties such as tensile strengths of the order of several hundred gigapascals, assuming a strain-to-failure of 30% and Young's modulus of about 1000 GPa. Yu *et al.* (2000a) have extended their experimental treatment to ropes of single-walled nanotubes stressed under *in situ* conditions inside a scanning electron microscope equipped with a custom-designed tensile stage. Although the scatter in the data was fairly significant, they concluded that most single-walled tubes failed at strain values of 5.3% or lower, and correspondingly their tensile strength was on the order of tens of gigapascals, with an average value of about 30 GPa. Their experiments also indicated that the Young's modulus of the single-walled tubes was typically on the order of several hundred gigapascals, with an average value of about 1000 GPa.

One of the key measurements of electronic properties involving single-walled carbon nanotubes was the electrical resistivity of ropes produced by laser ablation. This was achieved by the more precise approach of using a four-probe technique employed by Thess *et al.* (1996), utilizing multiwalled carbon nanotubes as the probes to measure the voltage drop across a short length of a rope of single-walled tubes. Their results indicated that the single-walled tubes were the most conductive form of carbon fiber ever,

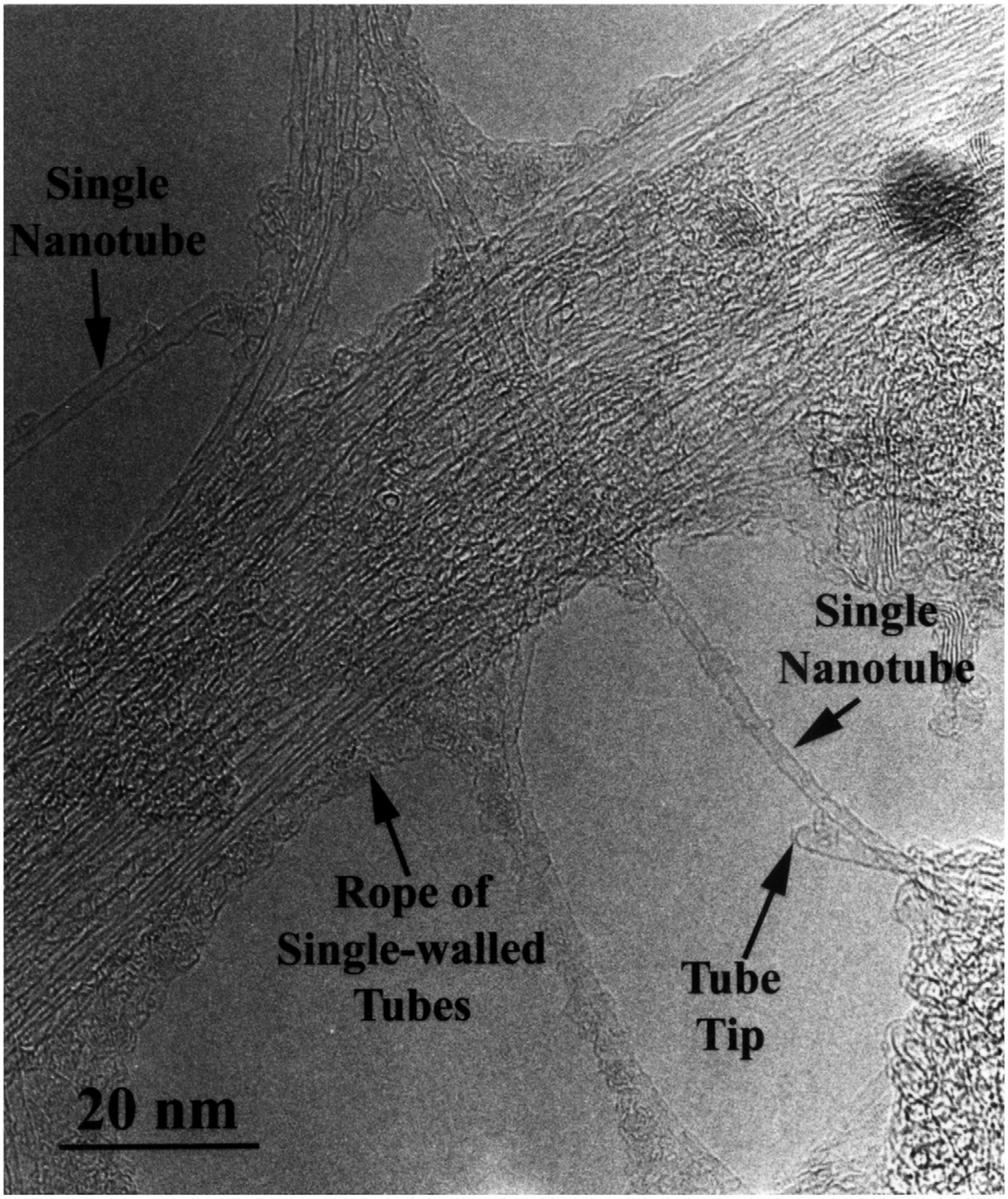


Figure 6
High-resolution TEM image of single-walled carbon nanotubes illustrating single tubes and ropes of several tubes.

with resistivities probably an order of magnitude lower than their multiwalled counterparts. The conductivity of single-walled nanotubes, in conjunction with their nanoscale dimensions, would render them as ideal quantum wires, where currents would increase or decrease in stepwise mode rather than in a continuous mode.

3. *Large-scale Synthesis of Carbon Nanotubes*

In the latter half of the 1990s, the euphoria of the discovery of various forms of nanocarbons such as the multiwalled and single-walled variants of carbon nanotubes and the development of a detailed understanding of their structure and properties was slowly

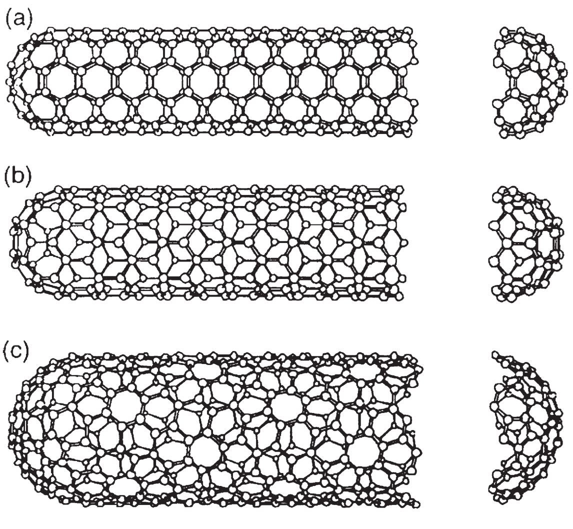


Figure 7
Schematic representation of different conformations of single-walled carbon nanotubes: (a) armchair, (b) zigzag, (c) chiral (reproduced by permission of Elsevier Science from *Carbon*, 1995, **33**, 883–91).

giving way to the need to develop real-life applications for these unique materials. Carbon nanotubes for research had been synthesized by the evaporation of carbon either in a DC arc-discharge or by laser ablation. Hand-in-hand with the need to explore real-life applications was the need to process them in substantially higher quantities and in specific configurations at costs commensurate with those of commercially available materials for specific applications, as well as the need to demonstrate specific applications where nanotubes performed better than any competing material. A very diverse portfolio of applications ranging from molecular composites to microelectronics to nanoprobes for chemical and/or biological applications to storage devices for fuels such as hydrogen were envisaged by Yakobson and Smalley (1997).

Carbon nanotubes are clearly very interesting materials considering their unique microstructure, their low density, and the fact that they possess extremely attractive mechanical and electronic properties. In order to harness them to exploit their full potential, the development of controlled synthesis methods to process nanotubes in bulk quantities and/or in ordered layouts are significant needs. In the electronics industry, materials in specific patterns and arrays are extremely critical as well. The major drawback of vaporization techniques such as DC arc-discharge and laser ablation is that both of them are extremely uncontrolled in terms of process parameters, resulting in tubes that contain significant fractions of unwanted material and that are difficult to manipulate and assemble in specific designs. Recently, a significant amount of research has been directed towards the development of techniques for the controlled architecture synthesis of both variants of carbon nanotubes, with the ultimate goal of gaining control over tube orientation, location, and individual tube characteristics.

The relatively old-fashioned technique of chemical vapor deposition (CVD) was the technique of choice for several investigators seeking to process nanotubes in significantly larger quantities than achievable by DC arc-discharge. The key parameters for the successful synthesis of carbon nanotubes via CVD are the catalyst type and size, and the role of the support on which the catalyst is placed. Li *et al.* (1996) reported the first documented synthesis of large quantities of aligned carbon nanotubes by CVD of acetylene in nitrogen gas over mesoporous silica containing iron nanoparticles at above 700 °C. Subsequently, Ren *et al.* (1998) demonstrated that aligned carbon nanotubes could be synthesized over areas up to several square centimeters using plasma-enhanced hot filament CVD with sputter-deposited nickel foils as catalysts over glass substrates at temperatures lower than 666 °C. In both cases, the aligned carbon nanotubes were determined by subsequent TEM to be of the multiwalled type. A scanning electron micrograph of the aligned multiwalled carbon nanotubes obtained by Ren *et al.* (1998) following plasma-enhanced CVD is shown in Fig. 8. The self-aligning of these carbon nanotubes which contain little to no second phase is ascribed to the van der Waals interaction between each tube and its nearest neighbors. Subsequently the same technique was modified to deposit well-patterned and regularly spaced arrays of carbon nanotubes on islands of transition metal catalytic films obtained by evaporating the desired metal catalyst itself in patterned arrays through a shadow mask.

While CVD appears to be a commonly used technique for the processing of carbon filaments and multiwalled tubes, it was only in 1998 that CVD was employed for the synthesis of ropes of single-walled tubes using methane as the feedstock and iron oxide nanoparticles as the catalyst (Kong *et al.* 1998). The synthesis of single-walled carbon nanotubes using well-controlled steps is a very important requirement for this technology, considering that their dimensions and electrical conductivity render single-walled nanotubes as quantum wires. The key to effectively processing single-walled tubes by CVD appears to be the very effective dispersion of the catalyst nanoparticles on the surface of a suitable high area support, such as alumina. Subsequent work by the same group demonstrated that the catalyst (iron/molybdenum compared to iron) and the support (hybrid alumina/silica compared to alumina) also played very significant roles in determining the yield of single-walled tubes produced by CVD. However, the major challenge with respect to the controlled synthesis of single-walled nanotubes appears to be obtaining

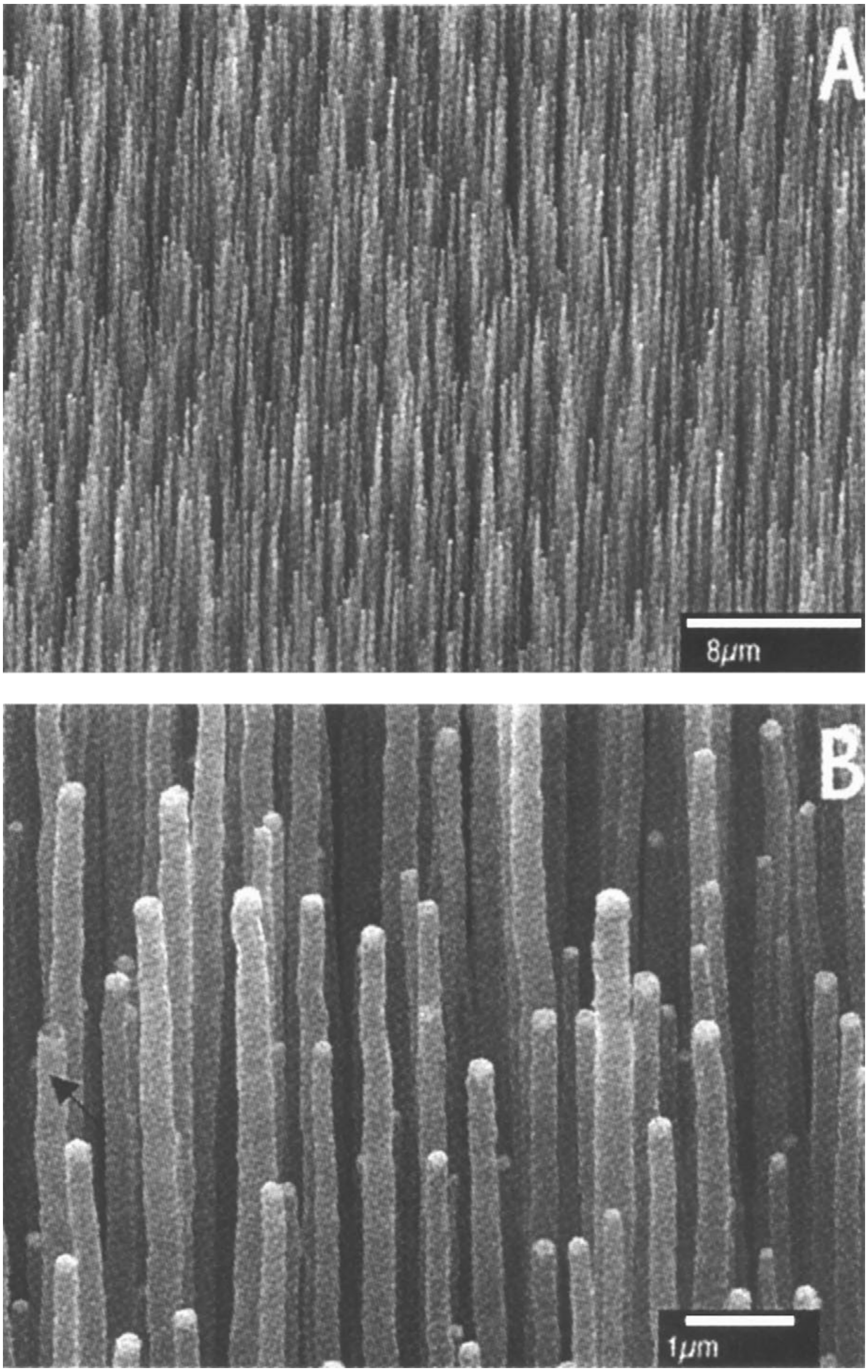


Figure 8
Scanning electron microscopy image of arrays of multiwalled carbon nanotubes grown on a glass substrate by plasma-enhanced hot filament CVD. The arrow in (b) indicates a nanotube from which the catalyst particle has been removed (reproduced by permission of the American Association for the Advancement of Science from *Science*, 1998, **282**, 1105–7).

specific aligned orientations. There appears to be substantial room for optimizing various conditions to overcome this and other potential obstacles. The ultimate goal with patterned growth of nanotubes is the direct fabrication of devices that would benefit immensely from the unique properties of these materials.

4. Applications and Potential Applications of Carbon Nanotubes

The unique combination of the physical structure (high aspect ratio and nanometer-scale radius of curvature of the tip), mechanical properties such as stiffness and strength, and the chemical stability of the graphene plane that constitutes its surface makes the carbon nanotube the consummate field-emitting device. This was initially demonstrated for a multiwalled carbon nanotube by Rinzler *et al.* (1995) using an isolated single multiwalled nanotube. Subsequently, several research groups have demonstrated that thin ropes of single-walled nanotubes are superb field-emitting sources that show higher emission properties (lower turn-on voltage and higher emissivity) than their multiwalled counterparts. The heart of such devices is the cathode, where the nanotubes serve as emission sources following appropriate fabrication techniques. Typically, single-walled tubes show better emission uniformity than multiwalled tubes. In prototype demonstrations using an 11.25 cm diode-based flat panel display with ropes of single-walled tubes serving as emitting sources, Choi *et al.* (1999) have shown that the device can be turned on at low voltages ($<1\,\mathrm{V\mu m^{-1}}$) and the brightness measured on the green phosphor (ZnS:Cu, Al) of the screen was of the order of $1800\,\mathrm{candela\,m^{-2}}$ at $4\,\mathrm{V\mu m^{-1}}$. The combination of unique properties possessed by carbon nanotubes also renders them as possible candidates for ultra-fine probes for techniques such as scanning probe microscopy. Dai *et al.* (1996) have demonstrated that multiwalled carbon nanotubes constitute well-defined tips for scanning probe microscopy when they are attached to the silicon cantilevers of conventional atomic force microscopes (AFM). Due to their flexibility, the MWNT tips were found to be resistant to damage from tip crashes, while their ultra fine diameters enable them to probe recesses in the specimen surfaces more effectively than conventional silicon tips. The excellent electrical conductivity of nanotubes could also be potentially exploited to utilize them for scanning tunneling microscopy (STM).

The dependence of the electrical properties of single-walled carbon nanotubes with the structure (specifically, the chirality) can be exploited for their use in some very important and unique applications. Several researchers have shown that single-walled tubes exhibit metallic conductivity either if they have the arm-chair like (m, m) configuration (Fig. 7) or if they have a chiral structure (m, n) where $m–n = 3X$ integer. Alternatively, they exhibit semiconducting characteristics when $m–n \neq 3X$ integer. The semiconducting type of single-walled carbon nanotubes is observed to be particularly suitable for the very important application of serving as a sensor for different gases. This is a key requirement for several environmental and chemical process control issues. Work by Kong *et al.* (2000) has demonstrated that the conductance of the semiconducting variety of single-walled carbon nanotubes can be significantly changed when exposed to gases such as NO_x and ammonia. Noticeable differences in the conductivity of mats (containing both semiconducting and metallic varieties) of single-walled carbon nanotubes were also observed during the course of these experiments. Other potential applications involving one or multiple single-walled carbon nanotubes with specific conductivities include nanoelectronic devices and molecular-scale mechanical actuators.

A very significant potential application with substantial commercial importance is the possible use of single-walled carbon nanotubes as a safe hydrogen storage medium for hydrogen-fueled transportation systems. Work by Dillon *et al.* (1997) through the use of temperature-programmed desorption measurements has shown that about 4 wt.% of hydrogen can be stored in the capillaries of the single-walled tubes at room temperature and under a pressure of 500 torr of hydrogen. This may be attributable to the quantum mechanical behavior exhibited by single-walled carbon nanotubes, behavior shared by no other form of carbon. In more recent experiments using single-walled carbon nanotubes doped with alkaline metals, the results appear to suggest substantially higher amounts of adsorbed hydrogen at elevated temperatures (200–400 °C) at ambient pressures. However, concerns such as whether the hydrogen is incorporated into the nanotubes in the molecular or protonated forms, or whether the gas can be released for use as fuels under conventional conditions, are issues that have not yet been addressed adequately.

5. Concluding Remarks

In summary, the discovery of the multiwalled and single-walled variants of the carbon nanotube, their structure and properties, and potential applications have been discussed in this chapter. In conclusion, it suffices to note that while several concepts and very exciting potential applications for carbon nanotubes have been demonstrated, significant amounts of further R&D are necessary for any of these discoveries to become commercial realities.

Bibliography

Bacon R 1960 Growth, structure, and properties of graphite whiskers. *J. Appl. Phys.* **31**, 283–90

Bethune D S, Kiang C H, de Vries M S, Gorman G, Savoy R, Vazquez J, Beyers R 1993 Cobalt-catalyzed growth of carbon nanotubes with single-atomic-layer walls. *Nature* **363**, 605–7

Choi W B, Chung D S, Kang J H, Kim H Y, Jin Y W, Han I T, Lee Y H, Jung J E, Lee N S, Park G S, Kim J M 1999 Fully sealed, high-brightness carbon nanotube field emission display. *Appl. Phys. Lett.* **75**, 3129–31

Dai H, Hafner J H, Rinzler A G, Colbert D T, Smalley R E 1996 Nanotubes as nanoprobes in scanning probe microscopy. *Nature* **384**, 147–50

Dillon A C, Jones K M, Bekkedahl T A, Kiang C H, Bethune D S, Heben M J 1997 Storage of hydrogen in single-walled carbon nanotubes. *Nature* **386**, 377–9

Dresselhaus M S, Dresselhaus G, Eklund P C 1995a *Science of Fullerenes and Carbon Nanotubes*. Academic Press, San Diego, CA

Dresselhaus M S, Dresselhaus G, Saito R 1995b Physics of carbon nanotubes. *Carbon* **33**, 883–91

Ebbesen T W, Ajayan P M 1992 Large-scale synthesis of carbon nanotubes. *Nature* **358**, 220–2

Harris P J F 1999 *Carbon Nanotubes and Related Structures*. Cambridge University Press, Cambridge

Iijima S 1991 Helical microtubules of graphitic carbon. *Nature* **354**, 56–8

Iijima S, Ichihashi T 1993 Single-shell carbon nanotubes of 1-nm diameter. *Nature* **363**, 603–5

Li W Z, Xie S S, Qian L X, Chang B H, Zou B S, Zhou W Y, Zhao R A, Wang G 1996 Large-scale synthesis of aligned carbon nanotubes. *Science* **274**, 1701–3

Kong J, Cassell A M, Dai H 1998 Chemical vapor deposition of methane for single-walled carbon nanotubes. *Chem. Phys. Lett.* **292**, 567–74

Kong J, Franklin N R, Zhou C, Chapline M G, Peng S, Cho K, Dai H 2000 Nanotube molecular wires as chemical sensors. *Science* **287**, 623–5

Krätschmer W, Lamb L D, Fostiropoulos K, Huffman D R 1990 Solid C_{60}: a new form of carbon. *Nature* **347**, 354–8

Kroto H W, Heath J R, O'Brien S C, Curl R F, Smalley R E 1985 C_{60}: buckminsterfullerene. *Nature* **318**, 162–3

Olk C H, Heremans J P 1994 Scanning tunneling spectroscopy of carbon nanotubes. *J. Mater. Res.* **9**, 259–62

Ren Z F, Huang Z P, Xu J W, Wang J H, Bush P, Siegal M P, Provencio P N 1998 Synthesis of well-aligned carbon nanotubes on glass. *Science* **282**, 1105–7

Rinzler A G, Hafner J H, Nikolaev P, Lou L, Kim S G, Tomanek D, Nordlander P, Colbert D T, Smalley R E 1995 Unraveling nanotubes: field emission from an atomic wire. *Science* **269**, 1550–3

Thess A, Lee R, Nikolaev P, Dai H, Petit P, Robert J, Xu C, Lee Y H, Kim S G, Rinzler A G, Colbert D T, Scuseria G E, Tomanek D, Fischer J E, Smalley R E 1996 Crystalline ropes of metallic carbon nanotubes. *Science* **273**, 483–7

Yakobson B I, Smalley R E 1997 Fullerene nanotubes:$C_{1,000,000}$ and beyond. *Am. Sci.* **85**, 324–37

Yu M-F, Files B S, Arepalli S, Ruoff R S 2000a Tensile loading of ropes of single wall carbon nanotubes and their mechanical properties. *Phys. Rev. Lett.* **84**, 5552–5

Yu M-F, Lourie O, Dyer M J, Moloni K, Kelly T F, Ruoff R S 2000bb Strength and breaking mechanism of multiwalled carbon nanotubes under tensile load. *Science* **287**, 637–40

S. Subramoney
DuPont Company, Wilmington, Delaware, USA

Cast Irons

Cast irons have played an important role in the development of the human species. They have been produced in various compositions for thousands of years. Most often they have been used in the as-cast form to satisfy structural and shape requirements. The mechanical and physical properties of cast irons have been enhanced through understanding of the fundamental relationships between microstructure (phases, microconstituents, and the distribution of those constituents) and the process variables of iron composition, heat treatment, and the introduction of significant additives in molten metal processing. The interested reader is referred to compilations of micrographs of all of the many different types of cast iron (Davis 1996, pp. 356–81). This article examines the relationships between microstructure, processing, and properties of graphitic cast irons. Abrasion-, corrosion-, and heat-resistant irons are specialty grades which are not discussed here. Information about these grades can be found in specialty handbooks (e.g., Davis 1996).

1. Metallurgy of Cast Iron

Cast irons, alloys of iron, carbon, and silicon, contain carbon as graphite (pure carbon), as carbide (Fe_3C), or in solid solution in austenite (austempered ductile iron (ADI), matrix austenite with 1.7–2.1% carbon). Cast irons represent the largest tonnage of cast-to-shape products produced worldwide, 46.8 million tonnes in 1997, not including that produced in the countries of the former Soviet Union; this compares to cast aluminum products at 6.3 million tonnes and cast steel at 5.5 million tonnes (Kanicki 1998). Table 1 lists compositions and distinctive microstructural features for common cast irons. Ferrite and pearlite are defined identically to steel, and ausferrite is a two-phase, high-carbon austenite matrix with embedded ferrite lathes.

The composition of cast irons is defined by the carbon equivalent, CE = % carbon + 1/3% silicon, where CE = 4.3 is the eutectic composition (Walton and Opar 1981). The usual composition, hypo- or hypereutectic, is given in Table 1 for each iron. Carbon and silicon largely determine the microstructure of cast products immediately after solidification. Figure 1 is a schematic vertical section through the stable and metastable Fe–C–Si phase diagrams (isopleth), where solid lines represent the stable iron–graphite phase diagram and dashed lines represent the metastable iron–cementite phase diagram.

2. Solidification of a Hypoeutectic Gray Iron Alloy With CE = 4.0

Solidification begins at the austenite liquidus, T_L, proeutectic austenite dendrites growing into the

Table 1
Cast iron composition ranges and microstructures.

Type	C range (wt.%)	Si range (wt.%)	CE range	Usual composition	Eutectic product	Graphite shape	Common RT matrix microconstituents
Gray	2.5–4.0	1.0–3.0	2.8–5	Hypo	γ + graphite	Platelets	Ferrite, pearlite
Ductile	3.0–4.0	1.8–2.8	3.6–4.9	Hyper	γ + graphite	Spheres	Ferrite, pearlite, ausferrite
Malleable	2.0–2.6	1.1–1.6	2.4–3.2	Hypo	γ + carbide	"Popcorn"	Ferrite, pearlite
White	1.8–3.6	0.5–1.9	2–4.3	Hypo	γ + carbide		Ferrite, pearlite
Compacted	3.0–4.0	1.8–2.8	3.6–4.9	Eutectic	γ + graphite	Rods	Ferrite, pearlite

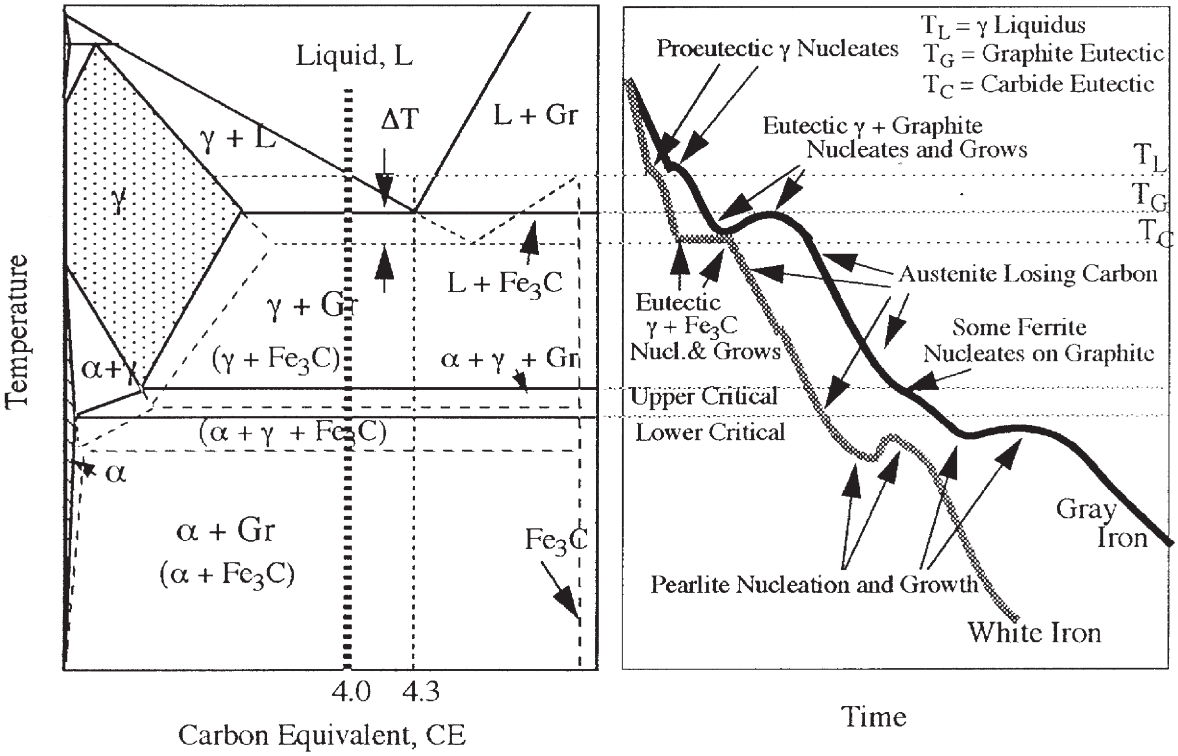

Liquid, L
ΔT
L + Gr
γ + L
γ
γ + Gr
(γ + Fe3C)
L + Fe3C
α + γ + Gr
α+γ
(α + γ + Fe3C)
α
α + Gr
(α + Fe3C)
Fe3C
Temperature
4.0
4.3
Carbon Equivalent, CE
TL = γ Liquidus
TG = Graphite Eutectic
TC = Carbide Eutectic
Proeutectic γ Nucleates
Eutectic γ + Graphite Nucleates and Grows
TL
TG
TC
Austenite Losing Carbon
Eutectic γ + Fe3C Nucl.& Grows
Some Ferrite Nucleates on Graphite
Upper Critical
Lower Critical
Gray Iron
Pearlite Nucleation and Growth
White Iron
Time

Figure 1
Schematic isopleth showing stable (graphite) and metastable equilibrium (Fe_3C) in Fe–C–Si alloys. Cooling curves for an alloy of CE = 4.0 are also shown.

liquid, enriching it in carbon until the stable eutectic reaction temperature, T_G, is reached (Merchant 1968). Cooling below T_G may result in the equilibrium eutectic reaction $L \rightarrow \gamma + Gr$, a microstructure consisting of austenite dendrites surrounded by a eutectic product (cells) of $\gamma + Gr$, a *gray iron* structure which would be represented by a cooling curve shown in Fig. 1. However, for higher cooling rates (see the more rapid cooling rate curve in Fig. 1), and/or a faulty processing history, the iron may supercool below T_G and satisfy metastable equilibrium between the phases α, γ, L and cementite, Fe_3C. The resulting structure would then be austenite dendrites surrounded by eutectic γ + Fe_3C, or *white iron.* The metastable phase diagram is shifted to lower temperatures than the stable diagram by an amount proportional to the amount of silicon. Adding 2% silicon increases the temperature difference, ΔT, between the stable and metastable eutectics from 4 °C to 30 °C. Similar effects are recorded in the eutectoid reaction range. These ΔT

values are exaggerated in Fig. 1 for illustration. Thus, each iron melted selects either stable or metastable transformation products during solidification. The tendency for iron to solidify white (defined as "chill") in practice is monitored with a "chill" wedge-shaped casting (a standard wedge block, ASTM A367, with thickness varying from zero to 25 mm (Davis 1996, pp. 45–6)), where the tendency to form white iron increases as the tip of the wedge casting is approached. Clearly, if the chill wedge has white iron up to and including the 1 cm width, then castings poured from the same iron with section sizes less than 1 cm would also be expected to have white iron present.

The possibility of white iron forming in the fast-cooling sections is cause for concern for foundries producing gray iron castings because of the negative effects of the hard iron carbide. One of the advantages of gray cast iron is its excellent machinability because of its graphite content, but if iron carbide is present it will quickly dull cutting tools. It is necessary to do everything possible to ensure that white iron does not form at any point in the gray iron casting.

Inoculation is a processing step where an addition (inoculant) is made to the melt just prior to pouring the castings to minimize the possibility of "chill." Inoculants usually contain reactive elements (calcium, aluminum, barium, strontium) in small amounts ($\sim 1\%$) which react with oxygen and sulfur to form the nuclei for graphite and therefore the graphite–austenite eutectic cells. The most direct result of inoculation is to reduce the amount of carbide in more rapidly cooling sections and to increase the number of eutectic cells (Lownie 1957).

3. *Matrix Microstructures in Graphitic Cast Irons—Cooling Below the Eutectic*

Matrix microstructures in graphitic irons are determined in a complex way by graphite size and shape, CE, section size, processing history, alloy content, and cooling rate. On cooling below the eutectic temperature (Fig. 1), austenite carbon content decreases until below the upper critical temperature, where austenite can begin to transform to stable ferrite + graphite. Ferrite will nucleate on graphite, depositing carbon from austenite that cannot be accepted by the b.c.c. ferrite (see the gray iron microstructure depicted in Fig. 2). In a slowly cooling casting, ferrite nuclei can continue to grow into the austenite. The final structure could be ferrite and graphite as predicted by the stable phase diagram. If the casting was cooled more rapidly, or in the presence of certain alloying elements, the austenite may transform to pearlite. This latter scenario is illustrated in Fig. 2 and in the cooling curve of Fig. 1.

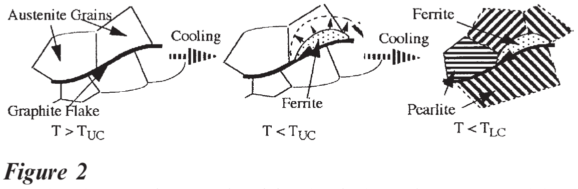


Figure 2
Nucleation and growth of ferrite below the upper critical temperature, T_{UC}, and pearlite below the lower critical temperature, T_{LC}.

Factors which favor the formation of ferrite include:

(i) high graphite surface to volume ratio—giving a high density of nucleation sites for ferrite;
(ii) high silicon content—silicon raises the upper critical temperature for $\gamma \rightarrow \alpha$, giving more rapid carbon diffusion and thereby making the growth of ferrite easier; and
(iii) slower cooling rates—annealing is used to produce ferritic structures.

Factors favoring the formation of pearlite would be the opposite. In addition certain alloy elements other than silicon would kinetically favor pearlite by lowering the equilibrium phase boundary temperatures.

4. *Microstructure and Mechanical Properties of Gray Cast Iron*

Gray cast irons have low tensile strengths and almost nonexistent ductilities (and impact strengths) due primarily to the nearly continuous nature (breaks at cell boundaries and at prior austenite dendrites) of the graphite flakes (see Fig. 3). Tensile stresses propagate cracks along the graphite plates internally. Tensile fracture stresses in gray cast irons are in the range 100–500 MPa (Schneidewind and McElwee 1950, Krause 1969). Attempts to determine the fracture toughness of gray irons by fracture mechanics methods yield conservative K_{IC} measurements in the range 16–22 MPa m$^{1/2}$ (Bradley and Srinivasan 1990).

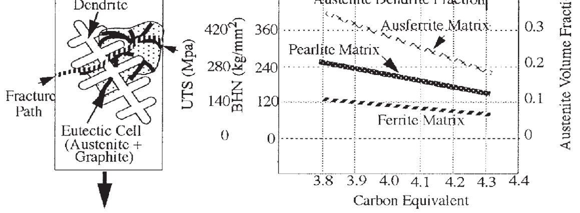


Figure 3
Schematic illustrating tensile stress fracture along graphite interfaces in gray iron. Schematic relationship is shown between UTS, BHN, and CE for different matrix microstructures.

5. Effect of Carbon Equivalent

The carbon equivalence of a gray cast iron affects the strength primarily because of its direct affect upon the amount of the proeutectic austenite dendrites present. This relationship (with a sketch illustrating a dendrite and the eutectic cell) and estimates of tensile strength are given in Fig. 3. The tensile strength of a eutectic iron (CE = 4.3) is typically around 160 MPa. As the amount of the proeutectic austenite increases with decreasing CE the tensile strength increases in the manner shown. The dendrites act as strengthening entities in much the same manner as adding stiffeners in a composite.

6. Effect of Matrix Microstructure

Ferritic irons have lesser strengths than pearlitic irons, which are in turn less than ausferritic irons (see Fig. 3). Clearly, the strength of pearlitic irons is greater because of the presence of the fine lamellar plates of iron carbide, which impede crack propagation, and the ausferritic matrix is stronger still because of an even finer dispersion of ferrite in a carbon-stabilized austenite matrix.

7. Effect of Alloying Elements

Alloying elements increase the strength of cast irons through their effect on the matrix. Common alloy elements include manganese, copper, nickel, molybdenum, and chromium, elements added primarily to control the matrix microstructure, having only a small effect upon the solidification microstructure. Two important mechanisms contribute to strengthening by alloying in the as-cast grades of gray iron, pearlite refinement (pearlite spacing gets smaller as a result of the alloy depressing the pearlite reaction to lower temperatures) and solid solution strengthening of the ferrite. As an example, an addition of 0.5% molybdenum to gray cast iron will increase the tensile strength by about 35–50 MPa for all values of CE. Transformation to pearlite or ferrite during cooling after casting is observed with cooling curves in much the same way as is done during solidification. A thermal arrest resulting from the latent heat given off during the transformation of austenite is shown in Fig. 1 for a gray iron with CE = 4.0.

8. Classes of Gray Cast Irons and Brinell Hardness

Gray irons are specified by their minimum tensile strength, i.e., a class 20 iron has a minimum tensile strength of 20 ksi (140 MPa), a class 30 iron has a minimum tensile strength of 30 ksi (210 MPa), etc. In most instances the chemical composition is not specified; rather it is the tensile strength and Brinell hardness which are written into specifications for certain applications. For example, automotive disk brake rotors and brake drums are produced from a class 30 gray iron with a Brinell hardness range of 179–229 kg mm^{-2}. It is usually true that alloy elements are not specified, only as needed to meet the minimum tensile specifications and the Brinell hardness range. The Brinell hardness of gray cast irons depends almost entirely upon the microstructure of the matrix material: ferrite ($\sim$100 kg mm^{-2}), pearlite ($\sim$250–kg mm^{-2}), and ausferrite ($\sim$250–350 kg mm^{-2}). Most specifications of matrix microstructures in gray irons call for pearlite, and it is highly desirable for the pearlite to be produced in the as-cast condition.

9. Ductile Cast Iron

Ductile cast iron was discovered in the late 1940s. In the period 1986–1997 the annual tonnage of ductile iron produced in the USA increased from 2.3 million tonnes to 4.1 million tonnes (Kanicki 1998). This growth is due to the microstructurally important fact that graphite in ductile iron is spheroidal, thereby making the iron-rich matrix constituent continuous. Cracks have no convenient weak path (as in flake graphite iron) through which to propagate, and the material takes on the properties of the ductile matrix, thus giving the name.

10. Production of Ductile Iron

Ductile iron (usually hypereutectic) is produced by "treatment" with magnesium immediately prior to pouring castings (see Fig. 4), followed by inoculation in much the same way as for the production of gray cast iron. Magnesium is added to the melt in the form of a master alloy ($\sim$46% silicon, 5% magnesium, balance iron). Dilution of magnesium reduces the violence of the subsequent reaction with molten iron. The castings must be poured within a short time after treatment, usually less than 10 minutes, because of the high vapor pressure of magnesium.

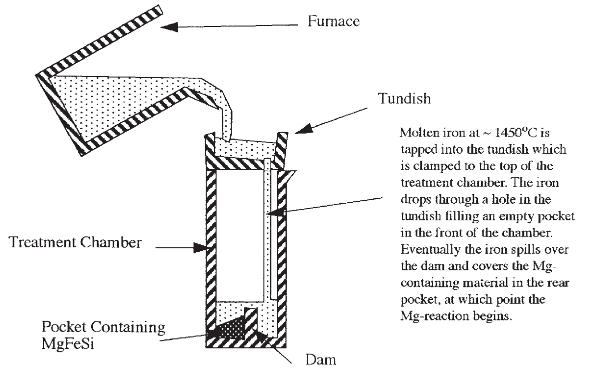
The reaction produces bubbles of magnesium vapor which rise through the molten iron bath. Successful treatment results when a significant portion of the magnesium dissolves in the molten iron (giving a composition of $\sim$0.03–0.05% magnesium), so that the correct conditions for graphite nodule formation are met in the solidifying melt. Magnesium contents $<$0.03% will result in vermicular or flake graphite, and contents $>$0.05% result in the appearance of exploded graphites. Shapes other than spherical contribute to a degradation in ductility. Typical "recoveries" of magnesium for Tundish treatment are in the range 50–60%. It is necessary that the sulfur content be kept $<$0.01% for successful treatment, because of the affinity of sulfur for magnesium (forming Mg_2S), thus removing elemental magnesium from the melt (Spengler 1982).

11. Solidification and Microstructures of Hypereutectic Ductile Cast Irons

Microstructural evolution during casting of magnesium-treated and properly inoculated irons with CE >4.3 is illustrated in Fig. 5. The events shown in Fig. 5 are listed below:

1 and 2. Nucleation and growth of graphite nodules, liquid composition following arrow A.

3. As the temperature drops below T_G, austenite nucleates on graphite enclosing the graphite in a shell. Simultaneously, conditions are met for the nucleation of austenite dendrites whose subsequent rejection of carbon moves the liquidus composition along arrow B.


Furnace
Tundish
Treatment Chamber
Pocket Containing MgFeSi
Dam

Molten iron at ~ 1450°C is tapped into the tundish which is clamped to the top of the treatment chamber. The iron drops through a hole in the tundish filling an empty pocket in the front of the chamber. Eventually the iron spills over the dam and covers the Mg-containing material in the rear pocket, at which point the Mg-reaction begins.

Figure 4
Schematic illustration of the Tundish method to produce ductile iron.

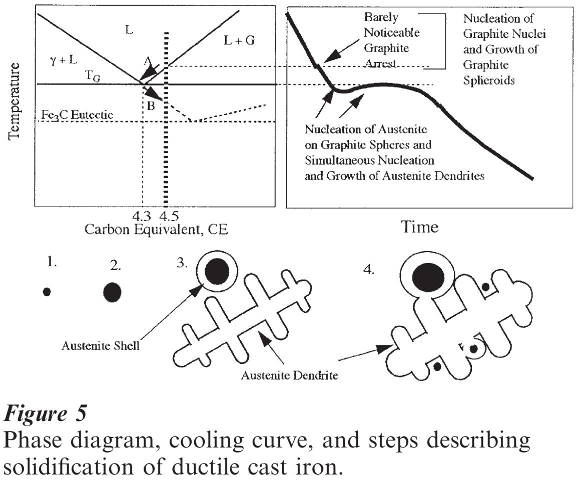

L
L + G
γ + L
T_G
A
B
Fe_3C Eutectic
Temperature
4.3 4.5
Carbon Equivalent, CE
Barely Noticeable Graphite Arrest
Nucleation of Graphite Nuclei and Growth of Graphite Spheroids
Nucleation of Austenite on Graphite Spheres and Simultaneous Nucleation and Growth of Austenite Dendrites
Time
1.
2.
3.
4.
Austenite Shell
Austenite Dendrite

Figure 5
Phase diagram, cooling curve, and steps describing solidification of ductile cast iron.

4. Supersaturation of carbon in the liquid adjacent to the dendrites can provide conditions allowing nucleation of new graphite nodules, which in turn grow and are enveloped in austenite.

This is a simplified view of a rather complex solidification event. This atypical eutectic reaction is called a neoeutectic, where growth of the graphite occurs by diffusion of carbon across the austenite shell into the graphite. The final as-cast microstructure depends upon the combination of section size, alloy content, and processing history. Measurements made on the plane of polish to characterize the graphite in ductile irons include determining the nodule count and the percentage nodularity.

A typical as-cast microstructure consists of a "bull's-eye" structure of ferrite surrounding graphite which in turn is surrounded by pearlite. The relative amount of ferrite present depends upon alloy content, nodule count, and the rate at which the casting is cooled through the transformation where ferrite and pearlite can form and grow, i.e., the same factors as those promoting ferrite in gray cast iron.

12. Mechanical Properties of Ductile Cast Iron

The mechanical properties of ductile cast iron depend almost completely upon the microstructure of the continuous matrix, a microstructure which can be controlled by heat treatment. An overview of the tensile properties (UTS vs. % elongation (EL) of ductile iron is given in Fig. 6. Impact properties of ductile irons are far superior to gray cast irons, again because of the continuous nature of the matrix. Upper shelf fracture toughness values of 90–100 MPa $m^{1/2}$ have been reported for ferritic

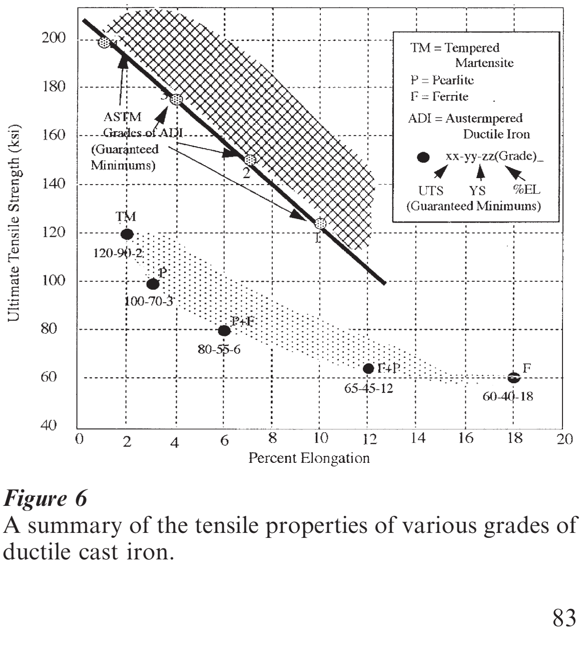


Figure 6
A summary of the tensile properties of various grades of ductile cast iron.

ductile irons and 60–70 MPa m$^{1/2}$ for pearlitic matrices (Bradley and Srinivasan 1990).

13. As-cast and Quenched and Tempered Grades of Ductile Iron

The UTS vs. % EL diagram of Fig. 6 illustrates the range of properties available in the as-cast and quenched and tempered conditions (Fuller *et al.* 1980, Davis 1996, pp. 63–76). Grade 60-40-18, with the highest percentage elongation, has a 100% ferritic matrix. Increasing pearlite means increasing strength, with a corresponding decrease in EL, until grade 100-70-3 contains 100% pearlite. The four as-cast grades shown are used in specifications as minimums for design purposes. The actual grade achieved will depend upon alloy content and section size. Grade 120-90-2 is heat treated where the iron has been austenitized, quenched to martensite, and tempered. Note in Table 2 that ferritic grades specify a maximum of the alloy elements copper and manganese, both of which promote the formation of pearlite when the casting is cooling. However, the grades with pearlite commonly have a specification with a minimum of copper. The elements manganese and chromium come from the steel component of the charge. Usually they are not present in quantities too large for the pearlitic grades, but it is becoming increasingly difficult to obtain steel with low enough "residuals" of manganese and chromium to allow the production of ferritic grades of iron in the as-cast condition. Many ferritic matrix castings are obtained by annealing, a heat treatment where the casting is austenitized and then slowly cooled, giving a large amount of time for the formation of ferrite. This heat treatment adds significantly to the cost of production and is not usually desired. Low residuals of manganese and chromium can be achieved by melting larger quantities of pig iron, a charge material which is usually more expensive than scrap steel.

Applications of ductile iron include numerous automotive and construction equipment castings; e.g., ferritic ductile cast iron is used to produce impact-resistant spindles (steering knuckles) for wheel applications, and pearlitic grades of ductile iron are a material of choice for automotive crankshafts.

14. Malleable Cast Iron, Processing, Microstructure, and Mechanical Properties

Malleable cast iron has lower CE (Table 1) than gray cast iron, solidifying white. The phase diagram and the cooling curve characteristic of white cast iron are illustrated in Fig. 1. It can be seen in the cooling curve that the temperature reached T_C before nucleation of graphite, thus resulting in white iron. The white iron castings are heated slowly into the $\gamma + Fe_3C$ two-phase field (Fig. 1) to a temperature in the range 800–1000 °C, and held for up to 24 hours allowing graphite (spheroidal "popcorn"-shaped) to nucleate and grow at the expense of the eutectic iron carbide. The final microstructure will be quite similar to ductile cast iron. Like ductile cast iron the matrix is the continuous phase and so the final structure will have similar properties to ductile cast iron, excellent strength, ductility, and toughness. The strength and ductility of malleable irons would overlap the as-cast properties of ductile iron shown in Fig. 6. Pearlitic malleable irons would be stronger but less ductile than ferritic malleable cast irons for the same reasons as for ductile iron. Fracture toughness values in the range 60–70 MPa m$^{1/2}$ have been reported (Bradley and Srinivasan 1990) for ferritic grades, values somewhat less than for a similar ferritic ductile iron. Ferritic malleable iron is produced by slow cooling through the upper critical temperature, T_{UC}, and pearlitic malleable iron by more rapid cooling, even cooling in air, after the malleablizing treatment (Gilbert 1968, Davis 1996, pp. 94–106).

Malleable cast iron has been produced for thousands of years, but it is slowly being replaced by ductile cast iron. Two important reasons why malleable iron is decreasing in favor is cost and limitations on section size that can be produced in the malleable state. Clearly it is less costly to produce nodules in the as-cast condition than to require an additional lengthy heat-treatment step. In addition ductile cast iron can be produced in very large section sizes, while

Table 2
Typical chemical composition (wt.%), matrix microstructure, and Brinell hardness.

Element	Grade 60-40-18	Grade 65-45-12	Grade 80-55-6	Grade 100-70-3	Grade 120-90-2
Carbon	3.5–3.9	3.5–3.9	3.5–3.9	3.5–3.8	3.5–3.8
Silicon	2.2–3.0	2.5–2.8	2.2–2.7	2.2–2.7	2.2–2.7
Manganese	0.30 max.	0.40 max.	0.2–0.5	0.6 max.	0.6 max.
Chromium	0.06 max.	0.10 max.	0.10 max.	0.10 max.	0.10 max.
Copper			0.2–0.4	0.2–0.5	0.2–0.5
Microstructure	Ferrite	Mostly ferrite	Mostly pearlite	Pearlite	Tempered martensite
Brinell hardness	130–170	150–220	170–250	241–300	270–550

the requirement to start with white iron limits malleable iron to practical section sizes of the order of 25–50 mm. Malleable iron section sizes larger than this will solidify with graphite flakes and would therefore have severely reduced ductility. Automotive applications for malleable iron continue to make large use of this material, including driveline yokes, connecting rods, and diesel pistons.

15. Compacted Graphite Iron

Compacted graphite (CG) irons are a form of cast iron recognized since the 1960s (Shellang 1965) as having unique mechanical and physical properties determined in large part by the shape of the graphite. These irons are produced in much the same way as ductile cast iron, except the objective is to have the molten iron dissolve a lesser amount of magnesium, usually ~0.01–0.015 wt.%. This amount of magnesium is enough to guarantee that no flakes are formed, yet not high enough to allow full nodularity (Davis 1996, pp. 80–93). A typical CG iron might contain 25% nodularity, the rest of the graphite having a twisted rod-shaped structure typical of CG iron.

The mechanical and physical properties of CG iron, for a given matrix microstructure, will reside between the properties of flake graphite iron and spheroidal graphite iron as shown schematically in Fig. 7. The more continuous nature of the compacted flakes will have physical properties closer to that of flakes (higher thermal conductivity and damping capacity), yet having tensile properties which are significantly superior to flake graphite irons.

16. Austempered Ductile Cast Iron

All as-cast and quenched and tempered grades of ductile iron have a b.c.c. matrix, with properties determined by the distribution of iron carbide within the ferrite matrix. Austempered ductile iron (ADI) has an austenitic matrix produced by heat treatment which is stable at room temperature (Johansson 1977). This austenitic matrix is responsible for the attainment of unique mechanical properties. Figure 6 shows ASTM grades of ADI having higher strengths than the as-cast grades, yet retaining excellent ductility, a result of the f.c.c. matrix together with a fine-scale dispersion of ferrite. The ASTM grades represent guaranteed minimums in tensile strength and ductility; the cross-hatched region at higher strengths and ductilities than the ASTM grades represents more typical ADI tensile properties. Clearly, ADI offers an entirely new opportunity for applications of ductile cast iron. Indeed ADI, with comparable strengths and ductilities, is making inroads into markets once dominated by steel forgings and steel castings.

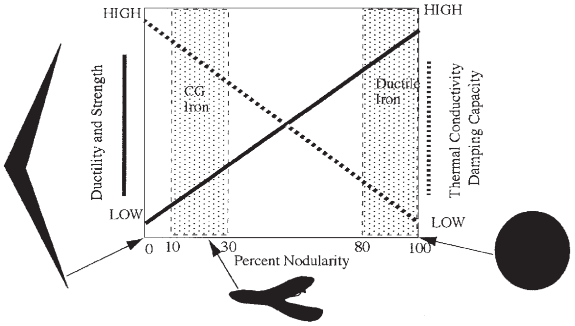


Figure 7
Variation of physical and mechanical properties of cast irons a function of the percentage nodularity of the iron.

17. The Metastable Phase Diagram and Stabilized Austenite

Consider the schematic low-temperature portion of the Fe–C–2.4 Si phase diagram in Fig. 8 together with an isothermal transformation diagram for a ductile cast iron austenitized at T_γ and austempered at T_A. Note that this diagram includes the martensite start temperature (M_S) and the metastable extension of the $(\alpha + \gamma)/\gamma$ phase boundary. The heat treatment used to produce ADI is illustrated in Fig. 8 and described as follows:

(i) Casting heated to and held at T_γ until fully austenitized with the matrix at 0.8 wt.% carbon.

(ii) Cooled to T_A, the austempering temperature, holding for 1–4 h, allowing the reaction γ (0.8 C) → α (0 C) + γ (2.0 C) (ausferrite).

(iii) Held at T_A until the ausferrite reaction is complete. Cooled to room temperature before the high-carbon austenite transforms to stable bainite.

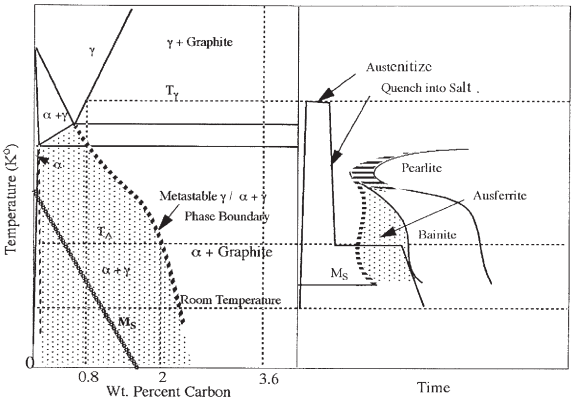


Figure 8
Isopleth at 2.4% silicon illustrating the metastable $\gamma/\alpha + \gamma$ phase boundary. Isothermal transformation diagram for ADI is also shown.

With the high carbon content present in the austenite, the martensite start temperature has been reduced to near absolute zero kelvin. This allows ADI to be austenitic at room temperature, an accomplishment made in commercial grades of cast iron only by adding large quantities of nickel or manganese (Davis 1996, pp. 192–204), or in steel by adding large quantities of nickel (austenitic stainless steel) or manganese (Hadfield steel).

Production of ADI to mechanical property specifications requires first producing quality ductile cast iron, and second, careful control of the heat treatment parameters necessary to convert austenite to ausferrite. Clearly, heat-treated products cannot overcome basic deficiencies in the cast state, including porosity (shrinkage and gas), slag, sand, inclusions from impurities, low nodularity, low nodule count, etc. These extrinsic defects can be minimized by good foundry practice and adhering to tight processing schedules. Some ‘‘deficiencies’’ cannot be overcome by good foundry practice which are intrinsic to the process of solidification. These include microsegregation in the casting and limitations in the nodule count that is attainable, a limitation defined primarily by the cooling rate (section size) of the casting. It has been shown that these intrinsic deficiencies can be minimized by proper selection of heat treatment and even a controlled amount of deformation at the austempering temperature before the ausferrite reaction commences (Rundman 1997).

18. Control of Mechanical Properties of ADI

ASTM grades of ADI illustrated in Fig. 6 can be created from the same alloy. The major factor controlling these tensile properties in ADI is the choice of the austempering temperature, the parameter primarily responsible for defining the scale (size and number of ferrite particles in the austenite matrix). High austempering temperatures (400–425 °C) give coarse microstructures (ASTM grade 1 ADI) with relatively low strengths but higher ductility, while low austempering temperatures (275–300 °C) give very fine microstructures (ASTM grades 4 and 5) with high strengths but lower ductility. Fracture toughness values for these materials give results in the range 60–90 $\mathrm{MPam}^{1/2}$ (Warda 1990).

Examples of potential applications for ADI include gears, automotive crankshafts, digger teeth for front-end loader applications, automotive camshafts, railroad car wheels, etc. Many applications require resistance to sliding wear, a property enhanced by the presence of the austenite which transforms to martensite in service.

19. Conclusion

This article briefly discusses the development of the microstructure and mechanical properties of cast irons containing graphite as a desired microconstituent. Phase diagrams and cooling curves are used to describe the fundamentals of solidification in gray and ductile cast irons. Important processing steps (inoculation in gray and ductile iron and magnesium treatment in ductile iron) required to produce quality gray and ductile cast irons are discussed. The mechanical properties (primarily tensile properties) of these irons are described in relationship to their microstructures. Malleable and CG irons are briefly noted and ADI is treated in detail.

See also: Austenite Decomposition, Overall Kinetics during Isothermal, and Continuous Cooling Transformation; Stainless Steels: Cast

Bibliography

Angus H T 1960 *Physical and Engineering Properties of Cast Iron*. BCIRA, p. 21

Bradley W L, Srinivasan M N 1990 Fracture and fracture toughness of cast irons. *Int. Mater. Rev.* **35** (3), 129–59

Davis J R (ed.) 1996 *ASM Specialty Handbook: Cast Irons*. ASM International, Materials Park, OH

Fuller A G, Emerson P J, Sergeant G F 1980 A report on the effect upon mechanical properties of variation in graphite form in irons having varying amounts of ferrite and pearlite in the matrix structure and the use of non-destructive tests in the assessments of mechanical properties of such irons. *AFS Trans.* **88**, 21–50

Gilbert G N J 1968 *Engineering Data on Malleable Cast Irons*. BCIRA

Johansson M 1977 Austenitic–bainitic ductite iron. *AFS Trans.* **85**, 117–22

Kanicki D P 1998 32nd census of world casting production. *Modern Casting* (Dec.), 54–55

Krause D E 1969 Gray iron—a unique engineering material. *Gray, Ductile, and Malleable Castings—Current Capabilities*, STP 485. ASTM, Philadelphia, pp. 3–28

Lownie H W 1957 Theories of gray cast iron solidification. *AFS Trans.* **65**, 340–51

Merchant H D 1968 Solidification of cast irons: a review of literature. In: Merchant H D (ed.) *Recent Research on Cast Iron*. Gordon and Breach

Rundman K B 1997 Segregation of alloying elements in ductile cast iron: the use of chemical, thermal, and mechanical processing to control the effects of segregation in austempered ductile iron. Honorary Cast Iron Lecture, American Foundrymen’s Society

Schneidewind R, McElwee R G 1950 Composition and properties of gray iron, parts I and II. *AFS Trans.* **58**, 312–30

Shellang R D 1965 Cast iron (with vermicular graphite) 3, 421–886 US Pat

Spengler A F (ed.) 1982 *The Ductile Iron Process, Compendium VI*. Miller, Chicago, IL

Walton C F, Opar T J 1981 *Iron Castings Handbook*. Iron Castings Society

Warda R D 1990 *Ductile Iron Data for Design Engineers*. QIT-Fer et Titane

K. B. Rundman
Michigan Technological University, Houghton,
Michigan, USA

Ceramic-modified High-temperature Polymers

Property enhancement through hybridization in the area of thermally stable polymers and ceramics is important for many applications in the electronics and aerospace industries. Preparation of such ceramers through the sol–gel process (Brinker 1988) allows systems to be designed that combine superior thermal stability, high refractive index, low thermal expansion coefficient, and a wide range of permittivity for ceramics, and to provide the toughness and crack deflection properties of polymers (Mark *et al.* 1995, Mark 1996). One of the challenges to obtaining significant improvements in these classes of polymers, however, is due to the fact that many of these polymers either possess superb thermal and mechanical properties to begin with or they are insoluble and infusible in organic solvents.

The unreactivity of such polymers that qualifies them as high-temperature, high-performance materials generally causes poor interfacial bonding between the polymers and the ceramic phases. As a consequence the molecular level combination of the two different components necessary to optimize the property combination may be difficult to achieve. Relatively few of the materials synthesized using these polymers have been exploited commercially. A brief account of the different strategies that have been used to combine thermally resistant polymers and ceramic phases using this process shall be discussed in this review.

The progress in heat-resistant polymers has been governed by the opposing requirements of thermal stability and processability. The structural requirement needed to attain high thermal stability means using the strongest chemical bonds in the polymer chain and resonance stabilization, and ensuring that these are no easy paths for rearrangement reactions. Polybonding is often utilized. Ring structures that have normal bond angles are preferred but to ease processability, thermally less stable flexible linking groups often have to be introduced between rings. Those having the least effect on stability are –COO–, –CONH–S–, $–SO_2–$, –O–, $–CF_2–$, and $–C(CF_3)_2–$.

Many thermally stable polymers, for example, polyesters, polyphosphazenes, aramids, and those containing heterocyclic rings: polyimides, polybenzoxazoles, polybenzobisthiazoles, and polybenzimidazoles have been used to combine with the inorganic components via the sol gel process. Silica is the ceramic phase most commonly generated in this sol–gel route to composites, but other ceramics such as titania (TiO_2), alumina (Al_2O_3), and zirconia (ZrO_2) have also been used. A combination of two or more oxides may also be used, as one type of oxide may improve one set of properties while the other improves different ones.

Morphology and phase separation control is critical in the generation of such hybrid materials. This problem has been solved by adding a coupling agent, functionalizing the oligomer or polymer chain ends, selecting polymers with appropriate backbone structures, or by using copolymers possessing functional groups that can interact with the growing inorganic oxide network. The relation between the structure of multiphase materials and their properties has been reviewed by many authors although there are severe limitations to theoretical models describing these parameters. In general the deformational behavior in such systems depends on the nature of the components, the volume fraction, size, shape, and orientation of the filler, and the state of adhesion between the filler and polymer matrix.

Radiation scattering, NMR, and other state-of-the-art techniques have been used to characterize these parameters. Generally these materials exhibit greater stiffness and hardness, although toughness is reduced. Highly transparent films, coatings and membranes have been produced using different means to control morphology. Preparation and properties of ceramers from some thermally resistant polymers are now described. Additional information is given in *Polymer–Ceramic Nanocomposites: Polymer Overview*.

1. Polyimides

The polyimides (PI) have become widely used as high performance polymers owing to their excellent thermal stability and toughness. However the thermal expansion coefficient (TCE), mechanical and dielectric properties, and so on, need to be strictly controlled for their application as circuit-printing films and semiconductor-coating materials in the microelectronics industry, and as adhesives in the aerospace industry. Several studies (Chujo 1996, Mascia 1995) have been carried out on the preparation of polyimide–metal oxide hybrid materials for this purpose through the sol–gel process. These materials have usually been produced by mixing solutions of pyromellitic dianhydride (PMDA), oxydianiline (ODA), and a metal oxide precursor, for example, tetraethoxysilane (TEOS), carrying out the hydrolysis and polycondensation of TEOS followed by an imidization process at high temperatures. Thin films were found to be transparent at low silica levels but were opaque at higher ones.

Despite the expected strong interactions through hydrogen bonds between the polyamic acid, the precursor for the PI, and hydroxyl groups from the silica phase, it was rather difficult to prevent phase separation. These systems have been compatiblized (Wang *et al.* 1994) with the selective use of thermally stable organically substituted alkoxysilanes, for example, aminophenyltrimethoxysilane, (aminoethylaminomethyl)-phenethyltrimethoxysilane, or 1-trimethoxysilyl-2-(*m*,*p*-chloromethyl)phenylethane, which undergo hydrolysis and polycondensation along with TEOS to form silica or polymeric silicates, while the groups at their other ends (such as the amino or chloro) provide secondary bonds with the PI chain.

Mascia and Kioul (1995) and Menoyo *et al.* (1998) have reported the use of glycidyl and isocyanate-type trialkoxysilanes to "couple" a polyamic acid to the silicate network prior to the condensation reactions for the formation of ceramer. By partial replacement of TEOS with dimethyldiethoxysilane and by varying the mixing conditions, morphologies ranging from interpenetrating networks to finely dispersed particles were obtained.

These results illustrate the role of the interconnected silica-rich particle within the polyimide-rich matrix in depressing the α-relaxations and reducing the TCE accordingly by an extent substantially greater than expected from the usual additivity rule.

Nandi *et al.* (1991) used the concept of "site isolation" to reduce the silica particle size. Constraining phase separation between the polymer and inorganic clusters to a smaller scale provides optically transparent materials with improved properties. Masa-aki *et al.* (1992) introduced functional groups into the polymer backbone using ethoxysilylated diamines in place of ODA which provided bonding sites for the silica. The silica particle size was found to decrease and the modulus to increase as the number of bonding sites increased. Habib (2000) partially replaced ODA with 2,2-bis(3-amino-4-hydroxyphenyl)methane, a diamine monomer that has two pendant hydroxyl groups to develop extensive bonding between the PI and the silica network. Transparent hybrid films containing 70 wt.% of silica were obtained.

The relationships between various mechanical properties and the silica content are given in Fig. 1 The tensile modulus increased with an increase in silica content. Unlike the unbonded system where the tensile strength decreases with an increase in silica level these ceramers show slight increases and then retention of tensile strength up-to much higher (25 .wt%) silica content. Extensive chemical bonding between the inorganic network and the polymer backbone solves the stress transfer problem thus reinforcing the hybrid films. The dynamical mechanical behavior of hybrid films also indicates (Fig. 2) the role of the bonded silica network in increasing the storage modulus and depressing substantially the α-relaxations. The SEM (scanning electron microscopy) micrograph of the unbonded ceramers (Fig. 3) shows sharp boundaries between the component phases, which become diffused in bonded ceramers (Fig. 4) showing interconnectivity between the phases. The thermal decomposition temperatures exhibited by these ceramers were at $\sim$550 °C.

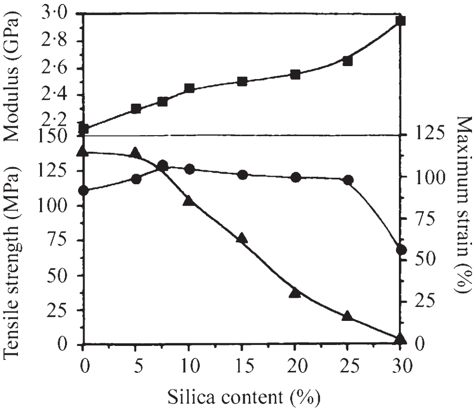


Figure 1
Comparison of the tensile properties as a function of the silica content in chemically bonded polyimide-silica ceramers. ● stress, ▲ strain

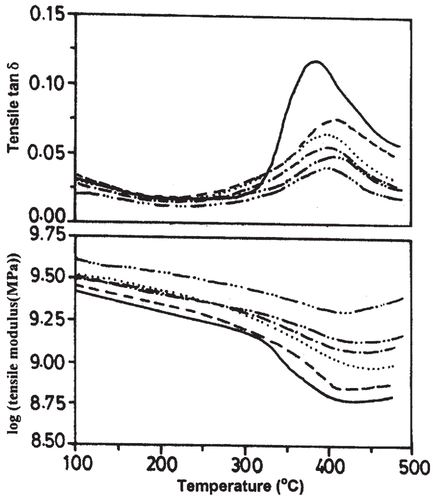


Figure 2
Temperature dependence of storage modulus and loss modulus for chemically bonded polyimide ceramers with silica wt%; ——, 0; —–, 10; · · · · ·, 15; –·–·, 20; –· ·–, 25; –· · ·–, 30.

Ree *et al.* (1995) prepared PI nanocomposite films for use in fabrication of microelectronic devices by solution blending of polyamic acid and silica aerogels, and subsequently processing the PI film by conventional methods. The optical and dielectric properties were improved, whereas the interfacial stress and thermal expansion coefficient were significantly degraded by a large disturbance in the polymer chain in-plane orientation caused by silica aerogels despite their low thermal expansivity. This indicates that in the rigid type of polymer composites,

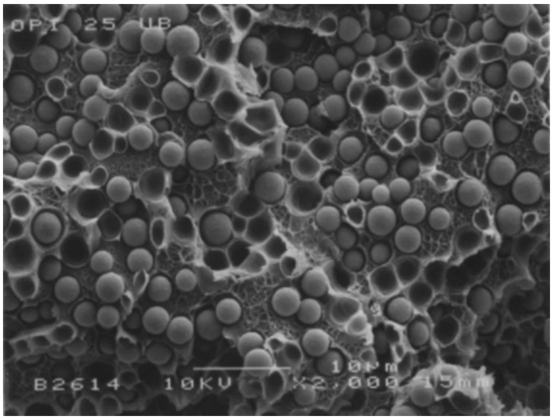


Figure 3
SEM micrograph showing discrete inorganic particles in nonbonded polyimide-silica ceramers.

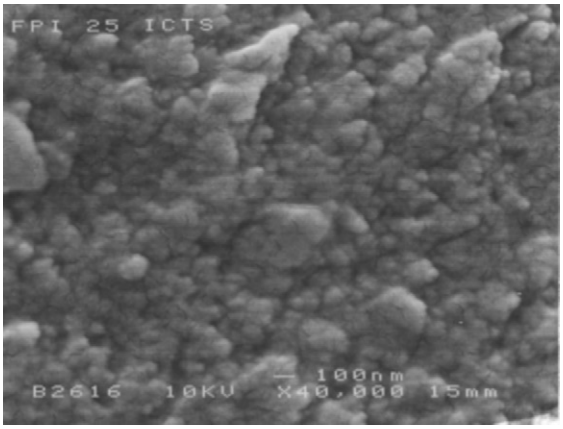


Figure 4
SEM micrograph showing diffused boundaries between the phases in chemically bonded polyimide-silica ceramers.

the orientation of polymer chains still has an important effect on the physical properties. Nano-composite membranes based on PI and titania, and silica components with a higher permeability coefficient and significant improvement in preselectivity, have also been prepared using the sol–gel process. The dielectric constant was lowered substantially.

The sorption and permeation data for gases such as O_2, CO_2, H_2 and CH_4 have been correlated (Hu *et al.* 1997) to the morphology of the composite structure. Transparent PI films with large nonlinear optical (NLO) coefficient moieties have also been prepared (Jeng *et al.* 1992) where the linked alkoxysilane dyes and the polyimide form a semi-interpenetrating polymer network resulting in a substantial increase in the α-relaxation temperature and superb thermal stability of the poled order. The poled film of this PI-inorganic composite showed a large second-order NLO effect and excellent temporal stability, over 168 h at 120 °C.

2. *Polyaramids*

Ahmad and co-workers, (Ahmad and Mark 1998, Ahmad *et al.* 1997a, 1997b, 1998) have synthesized various types of aramid–silica composites. These aramids have been prepared from a terephthaloyl chloride and a mixture of *para*- and *meta*-phenylene diamines, with the former acting to stiffen the chains and the latter to provide some flexibility and tractability. In those hybrid materials where the aramid chains were not chemically bonded to the silica network, no significant increase in the mechanical strength of the polymer was observed. In another system, end-capping the aramid chains with aminophenyltrimethoxysilane developed chemical bonding between the aramid chains and the silica network. By end-linking the polymer chains with the silica network, transparency of the films improved greatly. The tensile strength was found to go through a maximum, whereas the elongation at break decreased and the tensile modulus increased with increasing silica content.

Interconnectivity of the two phases is thought to be the reason for improvements in mechanical properties. Increases in silica content beyond 10 wt.% in the matrix, however, caused a decrease in the tensile strength. Interfacial bonding through end-linking of the polymer chain can be achieved only up to a limit in linear polymers, because the higher the molecular weight of the polymer, the fewer the chain ends and the more limited the bonding capabilities of the chain. In some related materials, the average functionality of the monomer was increased in the synthesis simultaneously of linear–nonlinear polyaramid chains, which could be bonded into the inorganic network at more than just the two ends. Such chains have been prepared by the addition of benzenetricarbonyl chloride and terephthaloyl chloride (TPC) in a two-step process in which the TPC was first reacted with the diamines, with addition of the trifunctional monomer delayed until near the end of the reaction.

The linear parts of the chains give the desired strong intermolecular bonding, while the nonlinear chain ends provide more bonding points and thus better adhesion with the inorganic network. The introduction of silica raises the modulus, but the toughness and elongation to break are greatly reduced. The bonding between such polymers and *in situ* generated silica significantly reduce water absorption, and this improves the mechanical and dielectric properties of the matrix, particularly under humid conditions.

3. Polybenzoxazoles

The successful use of pendant functional groups in benzoxazole copolymers to provide chemical bonding between the polymer chain and the silica network has been made to produce thermally stable transparent ceramers. The copolymers were prepared by the polycondensation reaction of 4,4′-(2,2,2-trifluoro-1-(trifluoromethyl)ethylidene)bis(2-aminophenol) with a mixture of 4,4′-dicarboxybiphenyl ether and 5 hydroxyisophthalic acid or 5-phosphonoisophthalic acid. The tensile strength was found to depend on the nature of copolymer and the number of segments containing pendant functional groups in the chain; it increased with initial increase in the silica content and then decreased. The mechanical strength of the copolymer containing phospho acid groups was lower than those containing the hydroxyl groups (Chen *et al.* 1995).

4. Polyphosphazenes

The sol–gel synthesis of hybrid materials based on poly[bis((methoxyethoxy)ethoxy)phosphazene] (MEEP) and silica has been reported. MEEP bears ethylene oxide oligomers as phosphorus side groups, which impart metal coordination ability to the polymer thus making it an ionic conductor. The low T_g of $-71\,^{\circ}\mathrm{C}$, has however limited the industrial development of MEEP. Chemical bonding of MEEP to a rigid inorganic network through the sol–gel process has been carried out (Guglielmi *et al.* 1996). Investigations on the thermal, mechanical, and electroconductive properties of the composite materials showed that these matrices presented improved mechanical and thermal features with respect to those of the original phosphazene. Abrasion-resistant transparent and antistatic MEEP–silica coatings have also been reported. Those prepared by the sol–gel process have been spin-coated on substrates and cured at ambient temperature or with moderate heating. The coating exhibits relatively steady conductivity with change in humidity, good transparency, toughness, hardness, and flexibility (Coltrain *et al.* 1990).

5. Polyesters

Transparent protective coatings from hydroxy-terminated polyesters and silica hybrids via the sol–gel process have been synthesized and evaluated for their performance on prefinish construction steel and aluminum. The interaction between the inorganic and organic domains occurs via hydroxyl end groups of polyesters and hydroxyl groups of silanols. These coatings combine flexibility, that is, upon deformation of the metal substrate after coating, with increased scratch resistance. The composition and molecular weight of the organic polymer, the ratio of the organic versus inorganic components, the cross-link density of the phases, solvents, catalysts, and curing conditions are some of the important parameters that control protein resistance of such coatings. Increases in Konig hardness and T_g was observed as the amount of inorganic component was increased (Frings *et al.* 1997).

6. Concluding Remarks

The inherent advantage of the sol–gel process and compatiblization of the organic and inorganic phases using organically substituted alkoxysilanes allows tailoring and improvement in materials properties. Many applications are still in the early stage of development, but transparent films, coatings, and membranes with high performance properties have been prepared and are being used in industry.

See also: High-temperature Stable Polymers

Bibliography

Ahmad Z, Mark J E 1998 Biomimetic materials: recent development in organic-inorganic hybrids. *Mater. Sci. Eng.* **C-6**, 183–96

Ahmad Z, Sarwar M I, Mark J E 1997a Chemically bonded silica–polymer composites from linear and branched polyamides in a sol–gel process. *J. Mater. Chem.* **7** (2), 259–63

Ahmad Z, Sarwar M I, Wang S, Mark J E 1997b Preparation and properties of hybrid organic–inorganic composites prepared from poly(phenylene terephthalmide) and titania. *Polymer* **38** (17), 4523–9

Ahmad Z, Sarwar M I, Mark J E 1998 Thermal and mechanical properties of aramid-based titania hybrid composites. *J. Appl. Poly. Sci.* **70**, 297–302

Brinker C J 1988 *Better Ceramic Through Chemistry*. Materials Research Society, Pittsburgh, PA

Chen J P, Ahmad Z, Wang S, Mark J E, Arnold F E 1995 Preparation and mechanical properties of benzoxazole–silica hybrid materials. In: Mark J E, Lee C, Bianconi P A (eds.) *Hybrid Organic–Inorganic Composites*. American Chemical Society, Washington, DC

Chujo Y 1996 Organic–inorganic hybrid materials. *Curr. Opin. Solid State Mater. Sci.* **1** (6), 806–11

Coltrain B K, Ferrar W T, Landry C J 1990 Abrasion resistant phosphazene ether–metal oxide composites and their preparation. Patent WO 9011323

Frings S, van der Linde R, Ten H, Meinema H 1997 Hybrid organic–inorganic coatings for prefinish construction steel and aluminium. In: Sarton L A J, Zeedijk H B (eds.) *Proc. Eur. Conf. Adv. Mater. Process Appl.*. Netherland Society for Material Science, Netherland, Vol. 3, pp. 267–70

Guglielmi M, Brusatin G, Facchin G, Gleria M, De Jager R, Musiani M 1996 Hybrid materials based on metal oxides and poly(organophoshazenes). *J. Inorg. Organomet. Polym.* **6** (3), 221–36

Habib R 2000 Study on the structure/property relationships in polyimide–silica hybrids with extensive interphase bonding prepared through sol-gel process. PhD Thesis, Saarbrucken Universitat, Germany

Hu Q, Marand E, Dhingra S, Fritsch D, Wen J, Wilkes G 1997 Polyamid-imide TiO_2 nano-composite gas seperation membranes: fabrication and characterization. *J. Membr. Sci.* **135** (1), 65–79

Jeng R J, Chen Y M, Jain A K, Kumar J, Tripathy S K 1992 Stable second order nonlinear optical polyimide-inorganic composite. *Chem. Mater.* **4**, 1141–4

Joly C, Goizet S, Schrotter J C, Sanchez J, Escoubes M 1997 Sol–gel polyimide–silica composite membrane: gas transport properties. *J. Membr. Sci.* **130** (1–2), 63–74

Mark J E 1996 The sol–gel route to inorganic–organic composites. *Heterogeneous Chem. Rev.* **3**, 307–26

Mark J E, Lee C Y -C, Bianconi P A (eds.) 1995 *Hybrid Organic–Inorganic Composites*. American Chemical Society, Washington, DC

Masa-aki K, Atsushi M, Yoshitake L, Yoshio I, Morikawa A, Iyoku I, Kakimoto M, Imai Y 1992 Preparation of new polyimide–silica hybrid materials via the sol–gel process. *J. Mater. Chem.* **2** (7), 679–90

Mascia L 1995 Development in organic–inorganic polymeric hybrids: ceramers. *Trends Polym. Sci.* **3** (2), 61–6

Mascia L, Kioul A 1995 Influence of siloxane composition and morphology on properties of polyimide-silica hybrids. *Polymer* **36** (19), 3649–59

Menoyo J D C, Mascia L, Shaw S J 1998 Compatibilization mechanism of polyimide–silica hybrids with organofunctional trialkoxysilanes. *Mater. Res. Soc. Symp. Proc.* Molecular-level ceramic polymer composites: synthesis of polymer-trapped silica and titania nanoclusters. **520**, 239–67

Nandi M, Conklin J A, Salvati H Jr, Sen A 1991 Molecular-level ceramic polymer composites: synthesis of polymer-trapped silica and titania nanoclusters. *Chem. Mater* **3**, 201–6

Ree M, Goh W H, Kim Y 1995 Thin films of organic polymer composites with inorganic aerogels as dielectric materials. Polymer chain orientation and properties. *Polym. Bull.* **35**, 215–22

Wang S, Ahmad Z, Mark J E 1994 Polyimide–silica hybrid materials having interfacial bonding through use of a sol–gel technique. *Macromol. Rep. A* **31**, 411–9

Z. Ahmad
Quaidi-i-Azam University, Islamabad, Pakistan

Chalcogenide Glasses

A chalcogenide glass is one containing a large amount of chalcogen atoms, sulfur, selenium, and tellurium. More widely, oxide glasses may be included in this category, but they are treated separately in the present article. A variety of stable glasses can be prepared in bulk, fiber, thin film, and multilayer forms using melt-quenching, vacuum deposition, and other less common techniques.

The atomic bonding structure is, in general, more rigid than that of organic polymers and more flexible than that of oxide glasses. Accordingly, the glass-transition temperatures and elastic properties lie in between those of these materials. Some chalcogenide glasses containing silver and so forth behave as (super-) ionic conductors.

These glasses also behave as semiconductors or, more strictly, they are a kind of amorphous semiconductor with bandgap energies of 1–3 eV. Electronic properties, which are governed by the species of chalcogen atoms, are modified by disordered structures, while the relationship between the structure and electronic properties remains to be studied.

Since several articles and texts have provided reviews on chalcogenide glasses, this section attempts to be an introduction to those publications, mainly emphasizing the nature of chalcogenide glasses in comparison with other materials. In addition, new insights are properly referred to wherever possible.

1. *Structural*

Glass formation of chalcogenide materials is relatively easy, and many kinds of glasses have been prepared by means of melt quenching, vacuum deposition, and other less common methods such as spin coating and mechanical amorphization (Borisova 1981; Elliott 1991; Feltz 1993; Kokorina 1996). Among the glasses, only selenium vitrifies as an elemental glass which is fairly stable at room temperature. Most of the stable binary glasses are compounds of a chalcogen atom and a group IV or V atom, such as Ge-Se and As-S, where atomic ratios can be varied widely. Ternary and more complicated glasses are of various kinds; these can probably be prepared by alloying with any kind of atom, including cations such as sodium and metallic atoms such as erbium, although the concentration may be limited. Oxychalcogenide and chalcohalide glasses can also be synthesized.

Thermal stability depends upon the glass. When heated, stable glasses such as As_2S_3 exhibit only the glass transition, some glasses such as selenium undergo glass transition and crystallization, and tellurium glasses are likely to crystallize directly.

As shown by the horizontal sequence in Fig. 1, chalcogenide glasses tend to possess properties intermediate between those of organic polymers and oxide glasses. Table 1 lists examples of elastic and thermal

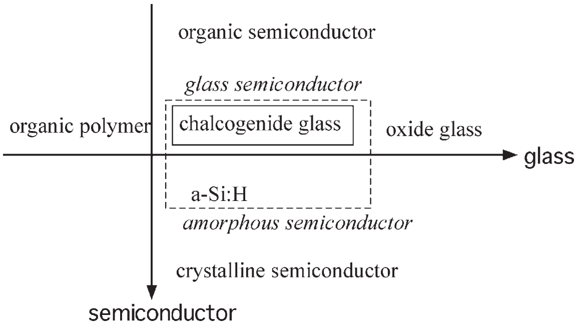


Figure 1
Characterization of chalcogenide glasses as glasses and semiconductors in comparison with other materials.

Table 1

Comparison of the density ρ, the bulk modulus B, the glass-transition temperature T_g, thermal conductivity κ, and the linear thermal-expansion coefficient β in polyethylene, As_2S_3, and SiO_2. The thermal conductivity of As_2S_3 differs among reports.

Material	ρ(gcm^{-3})	B (10^{10}Nm^{-2})	T_g (K)	κ (Wm^{-1}K^{-1})	β (10^{-5}K^{-1})
Polyethylene	0.95	~0.2	150	0.4	15
As_2S_3	3.2	1.25	460	0.2–0.5	2.5
SiO_2	2.2	3.65	1450	1.4	0.055

properties. On the other hand, chalcogenide glasses are denser than polymers and oxide glasses, since they are composed of heavier elements (S, Se, Te) than those (H, C, Si, O) in the others.

Chalcogenide glasses bear some similarity to oxide glasses, since both oxygen and chalcogens belong to group VI in the periodic table. The valence-electron configuration of all these elements is s^2p^4, and two *p*-electrons form atomic bonds with two neighboring atoms (Figs. 2 and 3). An exception is cation-chalcogen glass, which will be mentioned later.

However, in more detail, there exist marked differences between chalcogenide and oxide glasses, among which the most essential may be the ionicity of atomic bonds. For instance, according to Pauling's scale, As-Se and Si-O possess bond ionicities of 0.4 and 1.7. Taking the ionicity of Na-Cl of 2.1 into account, we can regard chalcogenide glasses as being covalent and oxides as ionic.

The difference in the ionicity causes different bonding angles (Elliott 1991, Feltz 1993). In the chalcogenide glass, for example in As_2Se_3, As-Se-As forms an angle of about 100°, which seems to be governed by covalent *p*-electron orbitals (Fig. 3). In contrast, in SiO_2 the Si-O-Si bond angle is distributed around 150°, which may be substantially influenced by Coulombic repulsive forces between positively charged silicon atoms. Hence, in As_2Se_3, $AsSe_3$ pyramidal units can be edge-shared (Fig. 2(b)), while in SiO_2 only corner-sharing of SiO_4 tetrahedral units is permitted (Fig. 2(c)). The coordination number, bond length, and bond angle specify the short-range structure up to the second-nearest neighbor atoms, which is nearly the same as that in the corresponding crystal.

The structure beyond the third-nearest neighbors, the so-called medium-range structure, remains controversial (Andriesh and Bertolotti 1997, Elliott 1991, Feltz 1993, Tanaka 1998). We may envisage distorted quasicrystalline structures (Fig. 2(b)), or otherwise, we may assume three-dimensional continuous random network structures, which were originally proposed for SiO_2 glass (Fig. 2(c)). In other words, the former emphasizes a structural similarity to the corresponding crystal, and the latter a universality in glass structures.

The structural controversy may be less serious for selenium (Elliott 1991, Feltz 1993). The material is known to be composed of entangled chains and/or ring molecules (Fig. 2(a)), depending upon the preparation conditions. Such structures are typical of organic polymers such as polyethylene and, because of this resemblance, amorphous selenium can be referred to as an "inorganic polymer."

In covalent chalcogenide glasses, as illustrated in Fig. 4, it has been argued that there exist some transitions at $\langle N \rangle = 2\cdot 4$ and $2\cdot 67$ (Elliott 1991, Phillips 1999). Here, $\langle N \rangle$ is the average coordination number of atoms, which is defined as $\langle N \rangle = \Sigma x_i N_i$, where x_i is the atomic fraction ($\Sigma x_i = 1$) of constituent elements and N_i is the atomic coordination number, which is 2 for chalcogens (Figs. 2 and 3).

Phillips and Thorpe have interpreted the transition at $2\cdot 4$ as follows (Phillips 1999): In a glass with $\langle N \rangle < 2\cdot 4$, the structure is floppy, and there exist so-called zero-frequency vibrational modes. That is, supposing only the bond length and the bond angle provide structural constraints, there remain free (zero-frequency) vibrational modes. At $\langle N \rangle = 2\cdot 4$, the structure becomes just rigid, and stable glasses can be prepared. Above this point, the structure is overconstrained, and glass formation becomes difficult.

However, the compositional dependence of some physical properties such as the density and the optical energy gap manifest extrema and/or thresholds also at $\langle N \rangle = 2\cdot 67$ (Phillips 1999). Tanaka assumes that this is a signature of a topological phase transition from two-dimensional ($<2\cdot 67$) to three-dimensional ($>2\cdot 67$) structures. As illustrated in Fig. 4, however, some physical properties such as the glass-transition temperature increase monotonically with $\langle N \rangle$. Further compositional studies are needed to obtain universal insights. It should be noted that such $\langle N \rangle$ dependence studies are inherent to covalent glasses, which can be varied continuously in *atomic* compositions.

On the other hand, with the vertical change in the periodic table from sulfur to selenium and tellurium, the atomic bond becomes more metallic (Borisova 1981, Feltz 1993). Directional bonding character changes to fairly isotropic in tellurium alloys, which are then likely to undergo crystallization relatively easily. The glass becomes darker in color, and the electrical conductivity becomes higher.

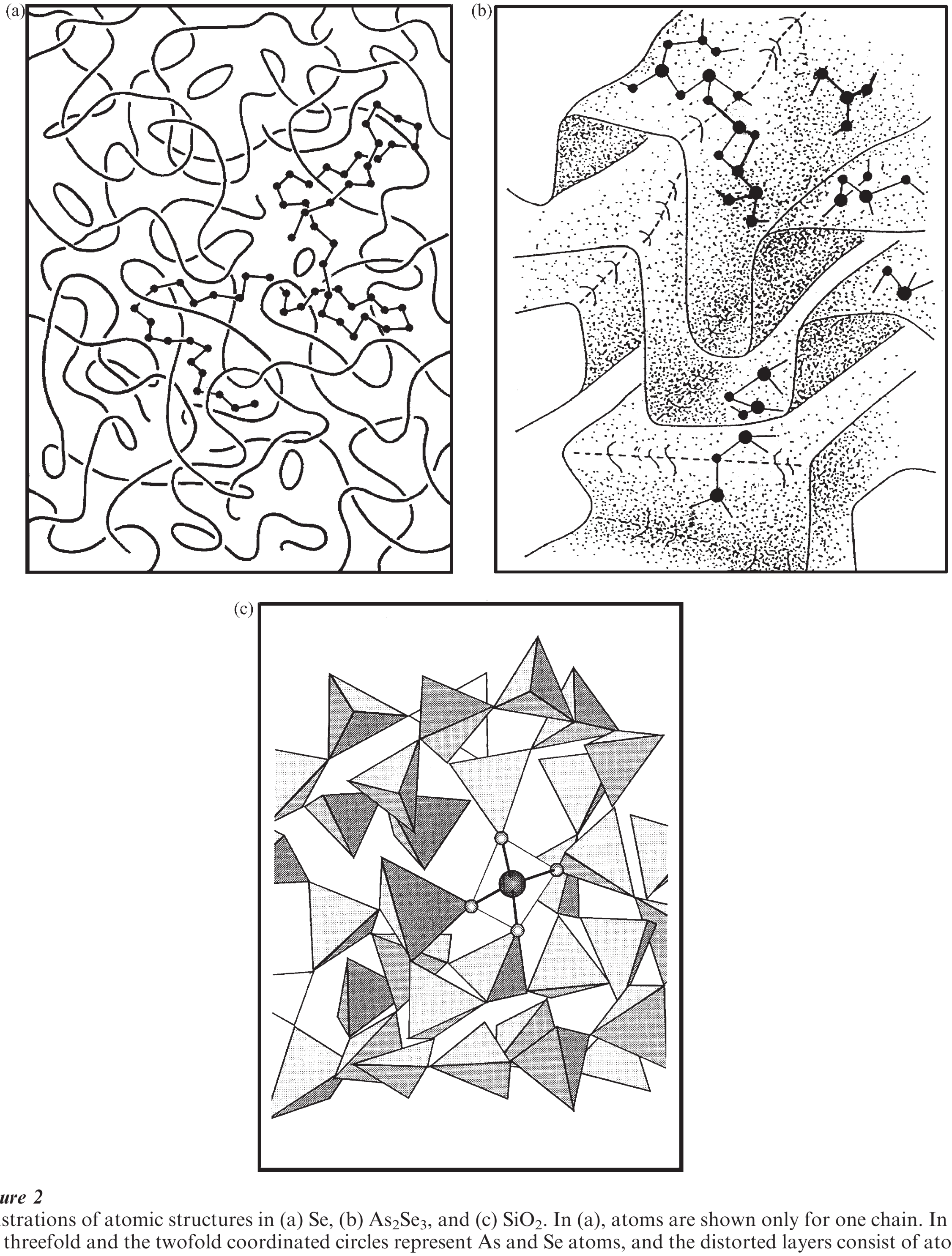


Figure 2
Illustrations of atomic structures in (a) Se, (b) As_2Se_3, and (c) SiO_2. In (a), atoms are shown only for one chain. In (b), the threefold and the twofold coordinated circles represent As and Se atoms, and the distorted layers consist of atomic clusters. In (c), the tetrahedra represent $SiO_{4/2}$ units. The bond length denoted by solid lines is typically 0·2 nm, and the interchain and interlayer distance in (a) and (b) is assumed to be 0.5 nm.

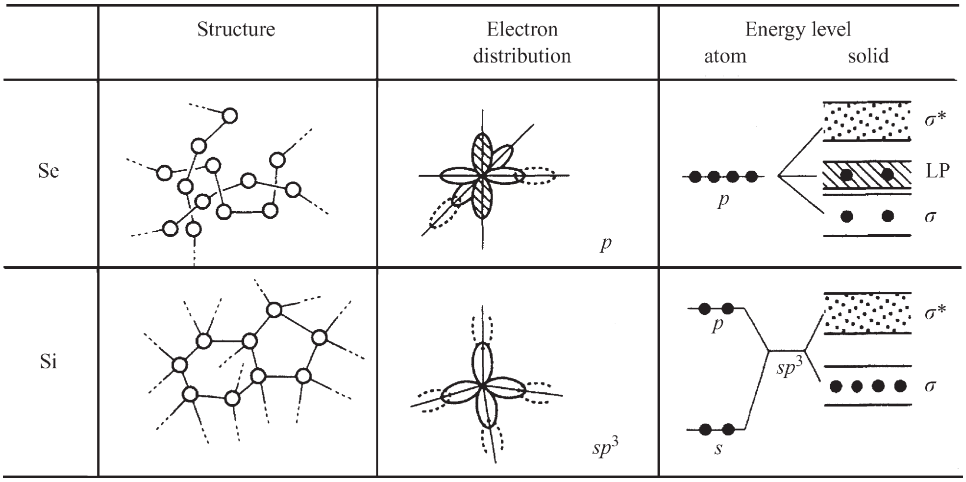


Figure 3
Comparison of a-Se and a-Si: the microscopic structure, the valence-electron distribution around an atom, and the electronic structures in an atom (left) and the solid (right).

Chalcogenide glasses containing group I–III atoms are exceptional from the above descriptions (Feltz 1993). These atoms are likely to behave as cations, which add some specific features. For instance, in Cu–As–Se, it has been demonstrated that the coordination number of selenium increases from 2 to 4. It is also known that in the Cu–As–Se system, glass formation is possible along the line connecting As_2Se_3 and Cu_2Se. This glass-formation region has been interpreted by Taylor *et al.* (1990) using a formal valence shell model, which assumes that lone-pair electrons are transferred to bonding electrons with compositional changes.

Another prominent feature, which should be pointed out here, is the ionic conduction in some glasses containing lithium, sodium, and silver (Borisova 1981, Feltz 1993, Kolobov and Elliott 1991). In these glasses, the ion forms atomic bonds with chalcogen atoms with relatively long bond distances of 0·25–0·3 nm. This feature and the high dielectric constant ($\varepsilon_r \approx 10$) in chalcogenide glasses favor marked ionic conductivity, since coulombic attractive forces are suppressed and the ions can move easily. Several glasses, such as Li_2S–P_2S_3–LiI, exhibit superionic conduction of $10^{-3}\,S\,cm^{-1}$ at room temperature. In contrast to the electronic conduction, which will be described in Sect. 2, the ionic conduction tends to be enhanced with structural disordering (Elliott 1991, Tanaka 1993). For instance, the Ag^+ conductivities in crystalline and glassy $AgAsS_2$ are reported to be 10^{-7} and $10^{-5}\,Scm^{-1}$.

Finally, it should be noted that the atomic structure and related properties in chalcogenide glasses depend upon preparation methods and history after preparation (Elliott 1991; Feltz 1993). This prehistory dependence is common to all nonequilibrium systems, such as glass. For instance, as shown in Fig. 5, Raman scattering spectra of As_2S_3 depend considerably upon the preparation methods.

2. *Electronic*

Chalcogenide glasses possess electrical and optical bandgaps of 1–3 eV and accordingly they can be regarded as amorphous semiconductors. The gap decreases in the sequence of sulfur, selenium, and tellurium, reflecting the enhanced metallic character. On the other hand, polyethylene and SiO_2 are insulators with energy gaps of 5–10 eV and accordingly these are transparent at visible wavelengths (Fig. 6).

As a semiconductor, the overall property of chalcogenide glasses can be grasped as the vertical sequence in Fig. 1. That is, with the change from organic semiconductors such as polyvinyl-carbazole, chalcogenides, hydrogenated amorphous silicon films, to crystalline semiconductors, the electronic mobility becomes higher and a faster response is available. The material also becomes more rigid. Instead, material prices seem to increase with this sequence, which may reflect their typical preparation methods, i.e., coating, evaporation, glow discharge, and crystal-growth techniques, respectively.

As exemplified in Fig. 3, the electronic structure of a chalcogenide glass is essentially the same as that of the corresponding crystal (Elliott 1991, Feltz 1993, Morigaki 1999). That is, in covalent glasses with the chalcogen coordination number of 2, the top of the valence band is composed of lone-pair *p*-electrons of chalcogen atoms, and the bottom of the conduction

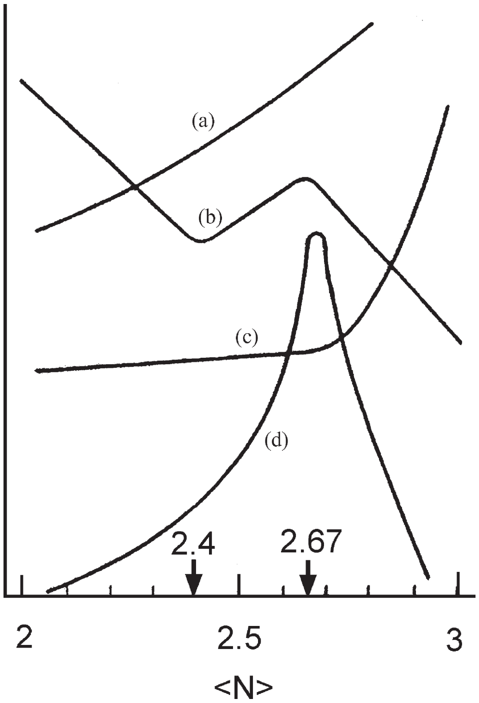


Figure 4
A schematic composition dependence of some properties as a function of the average coordination number $\langle N \rangle$. Examples are as follows: type (a) includes the position of the first-sharp X-ray diffraction peak and the glass-transition temperature; type (b), the bandgap energy, the specific volume, the specific heat, and the thermal expansion coefficient; type (c), the elastic constant and the density of tunneling states; and type (d), the intensity of the first sharp X-ray diffraction peak and the photodarkening effect.

band is composed of anti-bonding states of covalent bonds. The bandgap energy is similar ($\sim \pm 10\%$) to that in the corresponding crystal.

Due to this nature of the valence band, dramatic effects of pressure upon electronic properties appear (Elliott 1991, Tanaka 1998). Hydrostatic compression preferentially contracts the intermolecular distance of $\sim$0.5 nm, and then the overlapping between lone-pair *p*-electrons is enhanced. As a result, the valence band broadens, and the bandgap decreases remarkably.

Electrically, however, glasses exhibit smaller conductivities than the corresponding crystals (Borisova 1981, Elliott 1991, Feltz 1993, Tanaka 1993). This is because the electronic mobility is suppressed by band-tail and gap states, which are manifestations of disordered structures. Specifically, sulfide glasses such as As_2S_3 and GeS_2 behave as insulators, the conductivity being smaller than $10^{-15}\,S\,cm^{-1}$. On the other hand, the carrier density of intrinsic semiconductors is governed by the bandgap energy, and accordingly it is assumed to be similar to that in the corresponding crystal.

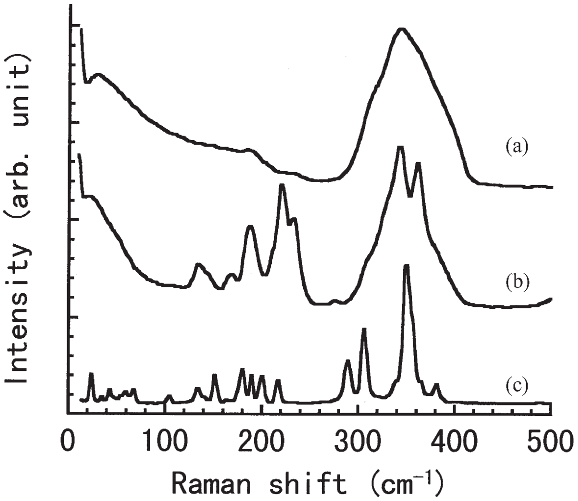


Figure 5
Raman scattering spectra of As_2S_3 in (a) a melt-quenched bulk glass and (b) an as-evaporated film. For comparison, a spectrum of the crystalline form is also shown in (c).

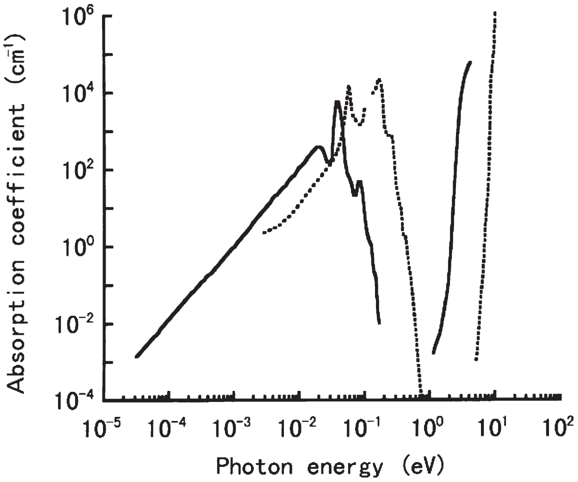


Figure 6
Optical absorption spectra of As_2S_3 (solid line) and SiO_2 (dotted line) glasses.

In detail, the glass can be regarded as a *p*-type semiconductor or, more exactly, hole conduction is greater than electron conduction (Elliott 1991, Feltz 1993, Madan and Shaw 1988, Morigaki 1999). The reason has not been elucidated. In selenium at room temperature, holes exhibit conventional Gaussian transport with a mobility of $0.1\,\mathrm{cm^2(Vs)^{-1}}$. However, at low temperatures and, in several materials, at room temperature, holes exhibit so-called dispersive transport, and the effective mobility decreases to $10^{-5}\,\mathrm{cm^2(Vs)^{-1}}$ or even less. Such hole motions are described in terms of multiple trapping with band-tail states or hopping transport in gap states. In general, the position of the Fermi energy, which may be located near the center of the bandgap, cannot be controlled by impurity doping.

However, there are a few exceptions, such as Bi–Ge–Se and Pb–In–Se, in which the thermopower indicates *n*-type conduction (Elliott 1991, Morigaki 1999). Hosono *et al.* (1998) have discovered a high *n*-type conductivity of $10^{-2}\,\mathrm{S\,cm^{-1}}$ in amorphous Cd–In–S films. In these films, sulfur atoms seem to be fourfold coordinated, as in crystalline CdS.

Optically, as shown in Fig. 6, chalcogenide glasses are transparent between red (or near IR) and IR regions (Borisova 1981, Elliott 1991, Feltz 1993, Morigaki 1999). The short-wavelength cutoff is determined by electronic excitation, and the long-wavelength limit by atomic vibrations; this feature is common in semiconductors and insulators.

The electronic absorption edge consists of three spectral curves (Elliott 1991, Feltz 1993, Madan and Shaw 1988, Morigaki 1999). In the absorption region higher than $\sim 10^3\,\mathrm{cm^{-1}}$, the absorption spectrum $\alpha(h\nu)$ in many materials follows the functional dependence $(\alpha h\nu)^{1/2} \propto h\nu - E_g^T$, and the optical bandgap energy is defined conventionally using E_g^T, the so-called Tauc gap. For smaller absorption coefficients, a temperature-independent Urbach tail with the form $\alpha \propto \exp(h\nu/E_U)$ is observed, where E_U is the Urbach energy. Selenium is exceptional here, for which E_U is markedly temperature dependent. Interestingly, for chalcogenides $E_U \gtrsim 50\,\mathrm{meV}$, which is comparable with $E_U \approx 50\,\mathrm{meV}$ in hydrogenated amorphous silicon. In the spectral region $\alpha \lesssim 10^0\,\mathrm{cm^{-1}}$, a weak absorption tail with a dependence of $\alpha \propto \exp(h\nu/E_W)$ is observed, where $E_W \approx 300\,\mathrm{meV}$ in As_2S_3. This absorption tail is influenced by impurities, but even in purified samples the tail still exists. It may be due to intrinsic mid-gap states. Interpretation of these spectral dependences still remains ambiguous.

Atomic structures responsible for the mid-gap states have not been experimentally confirmed, while the concepts, i.e., charged dangling bonds and/or valence alternation pairs, are frequently employed for the interpretation of many observations (Elliott 1991, Feltz 1993, Fritzsche 2000, Madan and Shaw 1988, Morigaki 1999, Shimakawa *et al.* 1995). The density of such native defects is estimated at 10^{15}–$10^{17}\,\mathrm{cm^{-3}}$ from considerations of the formation energy. For instance, the photoluminescence, which is excited by photons with $h\nu \approx E_g$ and which is emitted at $h\nu \approx E_g/2$, has been interpreted as a manifestation of structural transformations of these defects.

Photoelectrical properties of chalcogenide glasses have been studied extensively (Feltz 1993, Madan and Shaw 1988, Morigaki 1999), probably because of their unique photoconductive applications (see Sect. 4). Specifically, amorphous selenium films deposited onto substrates held at 50–60°C are highly photoconductive, and the characteristics have been investigated extensively. Among many features examined, the one that is markedly different from that in conventional crystalline semiconductors is the existence of a "nonphotoconducting spectral gap". As shown in Fig. 7, the photoconductive spectral edges in amorphous selenium are blue-shifted by more than 0.5 eV from the optical absorption edge. Geminate recombination of photoexcited carriers, which suppresses photocurrents, is assumed to be responsible for the spectral gap. In addition, avalanche multiplication of photocreated carriers has been demonstrated in amorphous selenium. This phenomenon appears to be anomalous, since the disordered structure is assumed to be unfavorable for producing accelerated carriers, which can cause impact generation of other carriers.

Chalcogenide glasses are known to exhibit prominent nonlinear optical effects (Andriesh and Bertolotti 1997, Elliott 1991, Kokorina 1996). Table 2

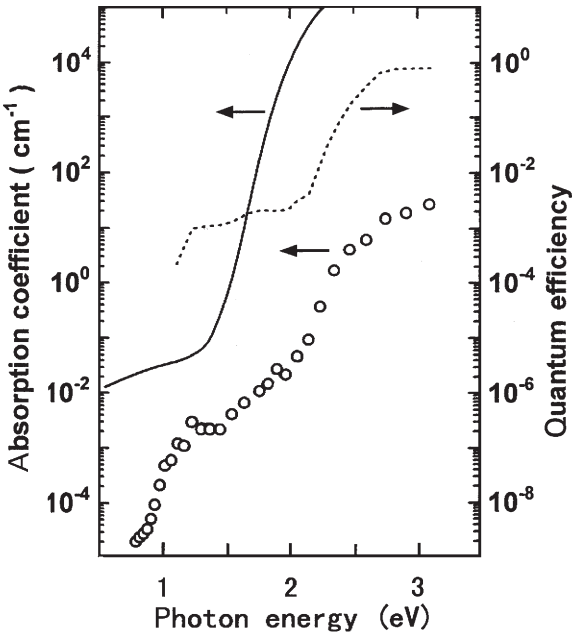


Figure 7
Comparison of some optical spectra for a-Se near the optical absorption edge. The solid line shows the optical absorption coefficient, the dotted line gives the photoconductive quantum efficiency, and the circles are the absorption coefficient derived from the constant-photocurrent method.

Table 2
Comparison of nonlinear coefficients in As_2S_3 and SiO_2 glasses. M_2 is the photoelastic figure of merit, M_a the nonlinear acoustic figure of merit, and $\chi^{(3)}$ the third-order nonlinear photoelectronic susceptibility.

Material	M_2 5 $(10^{15}s^{-3}\,kg^{-1})$	M_a $(10^{-12}s^2\,kg^{-1})$	$\chi^{(3)}$ $(10^{-14}esu)$
As_2S_3	433	240	500
SiO_2	1.5	0.4	3

Table 3
Photoinduced phenomena appearing in sulfide, selenide, and telluride glasses.

Material	Photon	Photo-thermal	Heat
Sulfide	darkening		
	fluidity	anisotropy	
Selenide		crystallization	
Telluride			phase-change

compares some nonlinear parameters of As_2S_3 and SiO_2 glasses. It can be seen that As_2S_3 is more efficient than SiO_2 by a factor of 10^2.

3. *Photoinduced Phenomena*

After the discovery of an optical phase-change phenomenon by Ovshinsky *et al.* in 1971 (Madan and Shaw 1988), extensive studies have been made on photoinduced phenomena in chalcogenide glasses (Andriesh and Bertolotti 1997, Elliott 1991, Fritzsche 2000, Shimakawa *et al.* 1995). As shown in Table 3, the phenomena can be grouped into three categories: the photon mode, in which the photoelectronic excitation directly induces atomic structural changes; the photothermal mode, in which photoelectronic excitation induces some structural changes with the aid of thermal activation; and the heat mode, in which the temperature rise induced by optical absorption is essential. Interestingly, these three kinds of phenomena are likely to appear in sulfides, selenides, and tellurides, respectively.

The best-known heat-mode phenomenon may be the optical phase change, or the so-called optical Ovonic effect (Fritzsche 2000, Madan and Shaw 1988). This phenomenon appears in tellurium compounds, which undergo thermal crystallization. A light pulse heats the film sample above the crystallization temperature and, as a result, a transformation from amorphous to crystalline phases occurs. In some materials, this change is reversible. That is, a more intense light pulse heats the sample above the melting temperature, and a successive temperature quenching can reproduce the original amorphous structure.

As for the photon-mode phenomena, reversible photodarkening and related changes have been extensively studied (Andriesh and Bertolotti 1997, Elliott 1991, Feltz 1993, Fritzsche 2000, Morigaki 1999, Shimakawa *et al.* 1995). Here, light illumination induces a red shift of the optical absorption edge, so that the sample becomes darker, while the red shift can be reversed with annealing at the glass-transition temperature. The refractive index in transparent wavelength regions increases with the red shift, which is consistent with the expectation obtained from the Kramers–Krönig relation. Sample volume, elastic, and chemical properties also change with illumination and recover with annealing. As for the light, bandgap illumination ($h\nu \approx E_g$) has been assumed to be effective. However, studies demonstrate that sub-gap light with the photon energy lying in the Urbach-tail region also produces some changes, which can be more prominent than those induced by bandgap light (Fritzsche 2000).

The photodarkening phenomenon continues to attract extensive interest, since this is a simple bulk phenomenon which is characteristic of the chalcogenide glass. That is, it does not appear in the corresponding crystal. Some structural studies have demonstrated that the amorphous structure becomes more disordered with illumination. However, it is difficult to identify explicitly the structural change in amorphous phases, and the mechanism is not yet elucidated.

An example of photothermal bulk phenomena is the photoinduced anisotropy originally discovered by Zhdanov and co-workers (Fritzsche 2000, Shimakawa *et al.* 1995). Macroscopically, untreated glasses are generally isotropic, while illumination with linearly polarized light can add some anisotropy such as dichroism, birefringence, and axial strains. Even unpolarized light can induce anisotropy if the light irradiates a side surface of a sample (Fritzsche 2000). In addition, the anisotropic direction can be altered by illumination with other polarized light. Furthermore, the anisotropy can be erased by illumination with unpolarized or circularly polarized light or by annealing. However, at lower temperatures, the photoinduced anisotropy appears less efficient and, accordingly, this phenomenon can be understood to be induced by a photothermal process. Fritzsche has assumed the process consists of directional changes of anisotropic structural elements, which exist in chalcogenide glasses. However, the atomic structure of the elements is controversial. It should be mentioned here that in some materials, such as amorphous selenium, oriented crystals can be produced by illumination with polarized light.

So-called photodoping, which was discovered by Kostyshin and co-workers, is a famous photothermal chemical reaction (Fig. 8) (Andriesh and Bertolotti 1997, Fritzsche 2000, Kolobov and Elliott 1999).

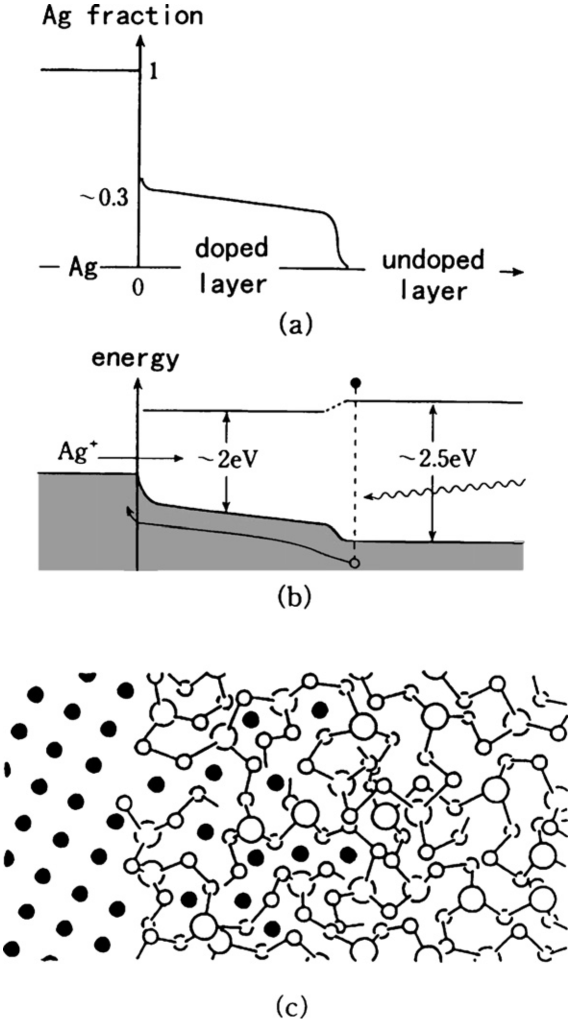


Figure 8
Schematic illustrations of a model for photodoping in an Ag:AsS_2 system when the photodoping is progressing. (a) shows the Ag atomic concentration, (b) the band diagram, where electron, hole, and Ag^+ motions are illustrated, and (c) the microscopic structure.

Consider a bilayer structure consisting of silver and AsS_2 films. When this bilayer is exposed to light, the silver film dissolves rapidly into the AsS_2. For instance, if the silver film is 10 nm in thickness, and the exposure is provided from the semitransparent silver side using a 100 W ultrahigh-pressure mercury lamp, the silver film will dissolve within a few minutes. Since the reaction becomes slower if the sample is illuminated at lower temperatures, this can be regarded as a photothermal phenomenon. However, it should be mentioned that temperature rise induced by illumination is not essential.

Much work has been done in order to understand the mechanism of this dramatic phenomenon. Owen has proposed that the reaction $Ag + AsS_2 \rightarrow AgAsS_2$ is favored thermodynamically and, accordingly, the composition of the photodoped layer becomes $AgAsS_2$. The driving force causing the silver dissolution seems to be provided by photoexcited hole motions. As for applications, photodoping has been demonstrated to be promising as a lithographic process, having an ultrahigh spatial resolution of about 10 nm.

In addition to the phenomena described here, there are also several irreversible and transitory changes. Examples are photopolymerization and photoinduced fluidity. The reader may refer to Fritzsche (2000) and Morigaki (1999).

Finally, it should be mentioned that not only photons but also other excitations such as electrons can give rise to a variety of structural changes (Andriesh and Bertolotti 1997). Some of these are similar to, but others are dissimilar from, the corresponding photon effects. For instance, electron beams can enhance silver doping into chalcogenide films like photons (Kolobov and Elliott 1991), while the beam can suppress crystallization of selenium, in contrast to the photocrystallization phenomenon. It should also be mentioned that studies using scanning tunneling microscopes, etc., demonstrate nanometer-scale structural changes in chalcogenide glasses (Utsugi 1990; Kado and Tohda 1997; Ohto and Tanaka 1998).

4. *Applications*

Four kinds of applications are commercially available or practically utilized. These rely upon the unique features of chalcogenide glasses: quasistability, photoconductive properties, infrared transparency, and ionic conduction.

The first is the phase-change phenomenon used in erasable high-density optical memories (Madan and Shaw 1988). The product utilizes semiconductor lasers and chalcogenide films such as GeSbTe with a thickness of ~15 nm. The reflectivity change between amorphous and crystalline phases is monitored with a weak light beam. The present memory capacity is about 5 GB/disk, which is still increasing (Mitsuhashi 1998).

The second category is photoconductive applications such as photoreceptors in copying machines and X-ray imaging plates (Elliott 1991, Feltz 1993, Madan and Shaw 1988). In a photoconductive target in vidicons, avalanche multiplication in amorphous selenium films is employed, which substantially enhances the light sensitivity so that star images can be taken.

The third application is purely optical (Elliott 1991, Feltz 1993, Kokorina 1996). That is, since the chalcogenide glass is transparent in IR regions, it can be utilized for IR optical components such as lenses and windows. It can also be utilized for IR-transmitting optical fibers. Fibers have also been employed as a matrix incorporating rare-earth ions, such as Er^{3+}. Such rare-earth ion doped chalcogenide glasses are promising for preparing functional fibers such as optical amplifiers.

Lastly, chalcogenide glasses containing group I elements such as silver are used as high-sensitivity ionic sensors (Elliott 1991, Feltz 1993). Some lithium-containing glasses can also be utilized as solid-state electrolytes in all-solid batteries.

5. *Final Remarks*

Chalcogenide glasses can be regarded as glassy semiconductors. For the atomic configuration, more-or-less firm insights have been obtained into the short-range structure, which covers the coordination number, the bond length, and the bond angle. Now, the problem is shifting to the identification of medium-range structures and defect structures. However, for these problems, there are no effective experimental techniques at present. Under such situations, we may employ computer simulation and/or refinement of simple structural models which can interpret as many observations as possible. In addition, studies of atomic composition dependence, which are practically possible only in covalent glasses, may add new insights into the chalcogenide-glass science. At the present stage, because of the incomplete structural knowledge, it is far behind the science constructed for single crystalline semiconductors.

Many problems, fundamental and simple, remain unresolved. For instance, the glass transition may be a common problem and also the biggest problem in glass science. On the other hand, electronic behaviors near the band edge and in the bandgap are also important problems both fundamentally and in applications. For instance, no one can give a convincing answer to such a simple question as "Why in chalcogenide glasses are holes more mobile than electrons?" Readers may refer to an issue of *Semiconductors* 1998, **32**(8), which discusses some problems.

Bibliography

Andriesh A, Bertolotti A (eds.) 1997 *Physics and Application of Non-Crystalline Semiconductors in Optoelectronics*. Kluwer, Dordrecht, The Netherlands

Borisova Z U 1981 *Glassy Semiconductors*. Plenum, New York

Elliott S R 1991 Chalcogenide glasses. In: Zarzycki J (ed.) *Materials Science and Technology*. VCH, New York, pp. 375–454

Feltz A 1993 *Amorphous Inorganic Materials and Glasses*. VCH, Weinheim, Germany

Fritzsche H 2000 Light induced effects in glasses. In: Boolchand P (ed.) *Insulating and Semiconducting Glasses*. World Scientific, Singapore, Chap. 10

Hosono H, Maeda H, Kameshima Y, Kawazoe H 1998 Novel *n*-type conducting amorphous chalcogenide $CdS \cdot In_2S_x$: an extension of working hypothesis for conducting amorphous oxides. *J. Non-Cryst. Solids*. **227–30**, 804–9

Kado H, Tohda T 1997 Nanometer-scale erasable recording using atomic force microscope on phase change media. *Jpn. J. Appl. Phys*. **36**, 523–5

Kokorina V F 1996 *Glasses for Infrared Optics*. CRC Press, Boca Raton, FL

Kolobov A, Elliott S R 1991 Photodoping of amorphous chalcogenides by metals. *Adv. Phys*. **40**, 625–84

Madan A, Shaw M P 1988 *The Physics and Applications of Amorphous Semiconductors*. Academic Press, Boston, MA

Mitsuhashi Y 1998 Optical storage: science and technology. *Jpn. J. Appl. Phys*. **37**, 2079–83

Morigaki K 1999 *Physics of Amorphous Semiconductors*. Imperial College Press, London

Ohto M, Tanaka K 1998 Scanning-tunneling-microscope modifications of Cu(Ag)-chalcogenide glasses. *J. Non-Cryst. Solids* **227–30**, 784–8

Phillips J C 1999 Constraint theory and stiffness percolation in network glasses. In: Thorpe M, Duxbury P M (eds.) *Rigidity Theory and Applications*. Kluwer, Dordrecht, The Netherlands, p.155

Shimakawa K, Kolobov A, Elliott S R 1995 Photoinduced effects and metastability in amorphous semiconductors and insulators. *Adv. Phys*. **44**, 475–588

Tanaka K 1993 Ion-conducting chalcogenide glasses: fundamentals and photoinduced phenomena. *Bulg. Chem. Commun*. **26**, 450–9

Tanaka K 1998 Medium-range structure in chalcogenide glasses. *Jpn. J. Appl. Phys*. **37**, 1747–53

Taylor P C, Saleh Z M, Liu J Z 1990 A general structural model for amorphous semiconductors. In: Fritzsche H (ed.) *Transport, Correlation and Structural Defects*. World Scientific, Singapore, pp. 3–25

Utsugi Y 1990 Nanometer-scale chemical modification using a scanning tunneling microscope. *Nature* **347**, 747–9

K. Tanaka
Hokkaido University, Sapporo, Japan

Chiral Smectic Liquid Crystals

1. *Chirality and Chiral Energies*

Molecules whose mirror images cannot be superposed are said to be chiral. Chirality is an absence of symmetry, and chiral molecules have a lower symmetry than do achiral ones. Since mirror images of linear or planar molecules can always be superposed, chiral molecules must necessarily have a three-dimensional structure. Typical chiral molecules have a spiral shape, or they have one or more chiral centers consisting of a carbon atom with four distinct units attached to its tetrahedral arms. Chirality gives rise to forces between molecules that favor relative twisting of their molecular axes, as depicted schematically for two spiral molecules in Fig. 1.

The desire of neighboring molecules to adopt configurations of relative twist gives rise to an astonishing variety of chiral liquid crystalline phases. These phases, which include the cholesteric phase, the blue phases, and the twist grain boundary (TGB) and SmC* phases to be discussed in this article, have new nonmolecular length scales set by the competition between chiral energies favoring twist and other energies favoring uniform parallel alignment of molecules. Even though chiral molecules must have a three-dimensional structure, they are effectively uniaxial, or nearly so, in most liquid crystalline phases because of spinning about their long axes. Thus, alignment of chiral liquid crystals can still be characterized by the Frank director **n**. The simplest chiral phase is the cholesteric, or chiral nematic (N*), phase depicted in Fig. 2. This phase is obtained by twisting the Frank director, **n**, of a nematic phase about an axis perpendicular to **n** to produce a twisted configuration,

$$\mathbf{n} = (0,\ \sin k_0 x,\ \cos k_0 x) \tag{1}$$

with pitch $P = 2\pi/k_0$ of the order of several thousand angstroms. The pitch wavenumber, k_0 (or equivalently the pitch P), in the cholesteric phase provides a convenient measure of the degree of chirality of chiral mesogens. The director of Eqn. (1) is characterized by a spatially uniform value of the twist:

$$\mathbf{n} \cdot (\nabla \times \mathbf{n}) = k_0 \tag{2}$$

and the spiral it forms is left handed when k_0 is positive. Twist, as defined in this equation, is a pseudoscalar: it changes sign under spatial inversion in which $\mathbf{n} \to -\mathbf{n}$ and $\nabla \to -\nabla$.

Chiral interactions give rise to a chiral contribution

$$f^*_{\mathrm{Ch}} = -h\mathbf{n} \cdot (\nabla \times \mathbf{n}) \tag{3}$$

to the Frank free energy, where h is a phenomenological parameter that depends in a complicated way on the degree of chirality of the constituent molecules and the short-range correlations of their spinning axes perpendicular to **n** (Harris *et al.* 1999). This energy, which is linear in the twist pseudoscalar $\mathbf{n} \cdot (\nabla \times \mathbf{n})$, clearly favors the development of nonzero twist. The full chiral Frank free energy

$$f^*_{\mathrm{n}} = f_{\mathrm{n}} + f^*_{\mathrm{Ch}} \tag{4}$$

where f_{n} is the Frank free energy is minimized by a cholesteric structure with twist $k_0 = h/K_2$, where K_2 is the Frank twist elastic constant.

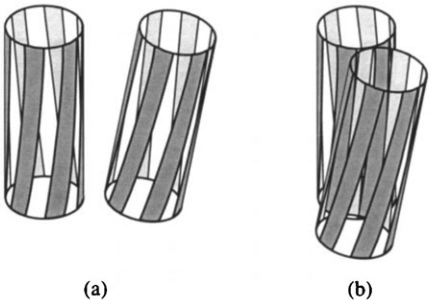


Figure 1
(a) Schematic of two chiral molecules in the form of a cylinder with a helical decoration. The darkened stripes are ridges and the white stripes are grooves. (b) The ridges in one molecule tend to align with the grooves of the second to produce the relative twist.

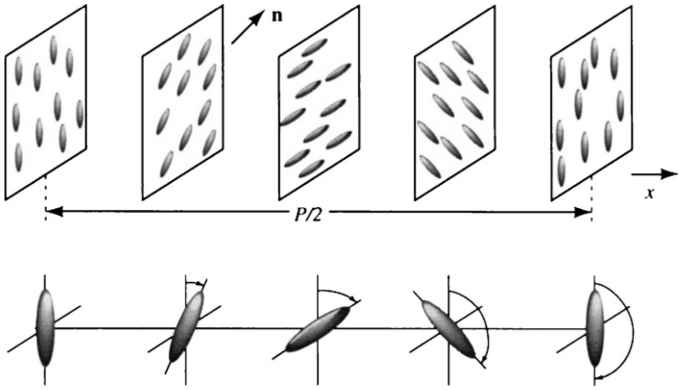


Figure 2
Schematic representation of the cholesteric phase. The Frank director **n** precesses in a helical fashion about a pitch axis perpendicular to local alignment planes. The director undergoes a rotation of π in a distance $P/2$ equal to half the pitch.

2. *Chiral SmA and TGB$_A$ Phases*

The layered structure of the SmA phase is fundamentally incompatible with the preferred cholesteric twisted structure of chiral molecules: it is impossible to twist the Frank director in an infinite smectic without destroying the layers in some way. Thus, there are two possible outcomes of introducing chiral molecules into a SmA phase: twist is altogether prevented by the smectic layer structure, or twist is allowed to enter the system by creating grain boundaries that disrupt the regular layer structure. In the former case, the result is a chiral smectic-A phase, denoted SmA*, that is structurally achiral but that exhibits typical properties of a chiral phase such as nonvanishing rotary power or optical dichroism. In the latter case, the result is a TGB$_A$ phase consisting of a periodic array of grain boundaries across which the Frank director and the smectic layer normal **N** undergo a rapid rotation through an angle $\Delta\theta$. This phase, whose experimental discovery (Goodby *et al.* 1988) coincided with the independent theoretical prediction (Renn and Lubensky 1988) of its existence, is depicted in Fig. 3.

A TGB consists of an array of parallel screw dislocations. A dislocation in a smectic liquid crystal is a topological line defect in which the smectic phase variable, u, undergoes a change of kd, where d is the layer spacing and k is an integer, in one circuit around a core. A screw dislocation, which has a linear core aligned along the layer normal, produces a helicoidal barber-pole-like layer structure shown in Fig. 4. The layers in a screw dislocation of strength k rise a distance of $kd/2$ in one half circuit around the core. Thus, as shown in Fig. 5, a periodic array of parallel unit-strength screw dislocations separated by a distance l_d produce a grain boundary across which the layer normal rotates by an angle

$$\Delta\theta = 2\sin^{-1}(d/2l_\mathrm{d}) \approx d/l_\mathrm{d} \tag{5}$$

where the final form applies for $d/l_\mathrm{d} \ll 1$. Thus, an array of N_b grain boundaries separated by a distance l_b will rotate layer normals through an angle of $N_\mathrm{b}\Delta\theta$ at an average rate of $\Delta\theta/l_\mathrm{b}$ per unit length. The chiral energy, f_n^*, is simply h times the average rate of change of the direction of the director **n**, which tracks the layer normal in smectic regions between grain boundaries. Thus, the chiral energy density associated with this grain boundary configuration

$$f_\mathrm{n}^* = -h\frac{\Delta\theta}{l_\mathrm{b}} \approx -h\frac{d}{l_\mathrm{d} l_\mathrm{b}} \tag{6}$$

is negative for $h > 0$.

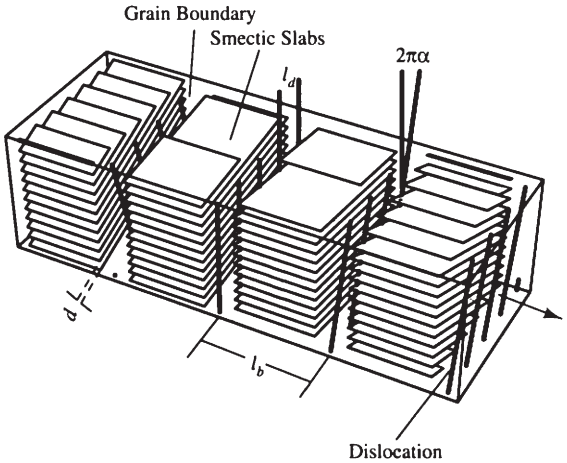


Figure 3
Structure of the TGB$_A$ phase. There is a periodic array of twist grain boundaries with separation l_b each consisting of parallel screw dislocations separated by l_d. The region between each adjacent pair of grain boundaries consists of a slab of parallel and flat smectic planes. The layer normals **N** in these slabs lie in the plane perpendicular to the pitch axis and rotate in discrete jumps across each grain boundary. In the TGB$_A$ phase, **n** is parallel to the layer normal in the smectic slabs. In the Renn–Lubensky TGB$_C$ phase, **n** makes an angle α relative to **N** in the plane perpendicular to the pitch axis.

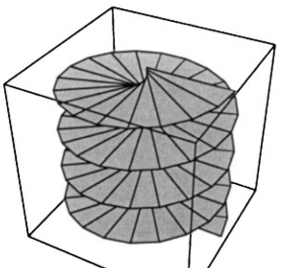


Figure 4
Screw dislocation.

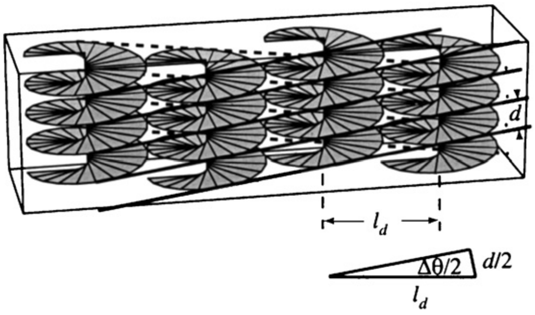


Figure 5
TGB depicted as an array of screw dislocations.

There is a one-to-one analogy between chiral smectic liquid crystals with twist and superconductors in an external magnetic field. The uniform SmA* phase with expelled twist is the analog of the Meissner phase in a superconductor (Tinkham 1996), and the TGB phase is the analog of an Abrikosov flux lattice (Abrikosov 1957, Tinkham 1996). As in superconductors, there are two types of smectic liquid crystals: type I systems that do not exhibit a TGB phase upon the introduction of chirality and type II systems that do. In type I systems, the ratio of the twist penetration depth, λ_2 (discussed more fully below), to the smectic correlation length, $\xi_\perp < 1/\sqrt{2}$. In type II systems, $\lambda_2/\xi_\perp > 1/\sqrt{2}$. Schematic phase diagrams in the temperature– chirality (T–h) plane for type I and type II systems are shown in Fig. 6. In type II systems, the TGB phase exists at a given temperature between an upper critical field h_{c2}, above which the TGB phase melts to a cholesteric phase, and a lower critical field h_{c1}, below which the smectic layer structure expels twist completely to form the Meissner-like SmA* phase. Since $\xi_\perp$ decreases with the phenomenological parameter $C_\perp$ in the NAC free energy, one expects type II behavior with a stable TGB phase to occur in the vicinity of the NAC point as shown schematically in Fig. 7.

The harmonic elastic free energy of a chiral smectic-A phase in terms of the phase variable u and the Frank director is

$$f^*_{\text{Sm}} = \frac{1}{2}Bu_{zz}^2 + \frac{1}{2}D(\nabla_\perp u + \delta\mathbf{n})^2 + f^*_{\text{n}} \tag{7}$$

where $\delta\mathbf{n} = \mathbf{n} - \mathbf{n}_0 \approx (\delta n_x, \delta n_y, 0)$. The combination $(\nabla_\perp u + \delta\mathbf{n})$ is invariant under simultaneous rotation of both the director and the layer normal. It arises from expanding the $(\mathbf{n}\cdot\mathbf{N})^2$ term in the energy of a smectic-A phase. If there are no dislocations present, the director relaxes to be locally parallel to the layer normal $\mathbf{N}$ so that $\delta\mathbf{n} = -\nabla_\perp u$, and the above free energy reduces to the familiar elastic energy for the SmA phase. Equation (7) defines the twist penetration depth: $\lambda_2 = \sqrt{K_2/D}$.

The energy per unit length of a single screw dislocation in the absence of the chiral energy f^*_{n} is easily calculated (Day *et al.* 1983) from Eqn. (7). The result is

$$\varepsilon = D\frac{d^2}{4\pi}\ln(\lambda_2/\xi_\perp) \tag{8}$$

Note this energy is finite like that of a superconducting vortex but not like that of a vortex in a superfluid. This is because the director can relax to screen singular configurations of the displacement field u.

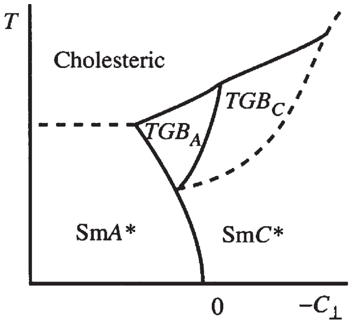


Figure 7
NAC phase diagrams in chiral systems. The region around the NAC point in achiral systems opens up into a series of TGB phases. Only TGB_A and TGB_C phases of an unspecified type are shown. First-order transitions are indicated by dotted lines and second-order transitions by full lines.

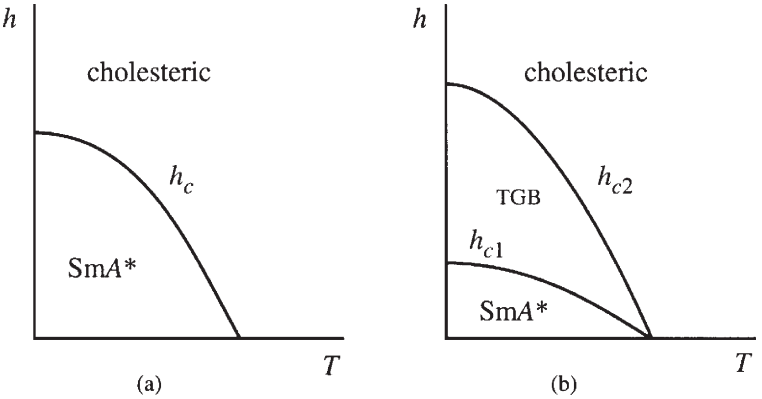


Figure 6
Schematic phase diagram for chiral smectics in the h–T plane (a) for type I systems in which there are only the cholesteric phase and the chiral SmA* phase with no twist and (b) for type II systems in which the TGB phase intervenes between the cholesteric and the SmA* phase.

As discussed above, a grain boundary consists of an array of dislocations. If interactions between dislocations are neglected, as they can be if their separation is sufficiently large, then the energy of N_b grain boundaries of side L, each containing N_d dislocations, is $E_{GB} = N_b N_d \varepsilon L$. This is a positive energy that opposes the introduction of twist. The chiral term, however, favors twist as is evident from Eqn. (6). The total free energy density of an array of grain boundaries is thus

$$f_{TGB} = \frac{1}{l_b l_d}(\varepsilon - hd) \tag{9}$$

This energy is positive for h less than the lower critical field

$$h_{c1} = \varepsilon / d \tag{10}$$

and negative for h greater than h_{c1}. Thus, for $h < h_{c1}$, it is energetically unfavorable for twist to enter the system, and the preferred state is the uniform smectic state with no twist. For $h > h_{c1}$, it is energetically favorable for the smectic to twist to form the TGB_A phase depicted in Fig. 3. The actual separation between dislocations and grain boundaries for $h > h_{c1}$ is determined by interactions between dislocations, which are complex but always repulsive.

3. Chiral Smectic-C Phases

In the SmC phase, molecules tilt on average relative to the smectic layer normal making it possible for the Frank director to develop a nonvanishing twist in two different ways. The twist pitch axis can either lie in the plane of the layers to produce a TGB_C phase (Renn and Lubensky 1992, Navailles *et al.* 1993), or it can lie along the layer normal to produce the chiral smectic-C (SmC*) phase (Meyer *et al.* 1975) in which the **c**-director (the projection of **n** onto the smectic planes) precesses in a helical fashion as shown in Fig. 8.

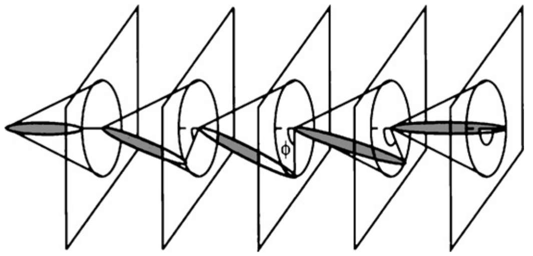


Figure 8
Schematic representation of the director in the SmC* phase. It has a constant component cos α along the layer normals. The **c**-director precesses in a spiral pattern in the plane of the layers with the pitch axis along the layer normals.

3.1 The TGB_C Phases

There are at least two possible versions of the TGB_C phase, one of which has received unambiguous experimental confirmation (Navailles *et al.* 1993). The conceptually simpler one (Renn and Lubensky 1992), sometimes referred to as the Renn–Lubensky phase, has a grain boundary structure identical to that of the TGB_A phase shown in Fig. 3 with layer normal always in the plane perpendicular to the pitch axis. In this phase, the Frank director also lies in the plane perpendicular to the pitch axis making an angle α with the layer normal as it does in a nonchiral SmC. In the second TGB_C phase (Navailles *et al.* 1993), sometimes referred to as the Bordeaux TGB_C phase, the director lies approximately in the plane perpendicular to the pitch axis, but the layer normal rotates in a cone with opening angle $(\pi/2) - \alpha$ relative to the layer normal as shown in Fig. 9. Since the layer normal is no longer perpendicular to the pitch axis, the grain boundaries comprising the phase have an edge as well as a twist character as can be seen in Fig. 9.

3.2 The SmC* Phase

To see how chirality gives rise to the SmC* structure shown in Fig. 8, it is sufficient to use the SmC parameterization of the director in f_n^*. The result is

$$f_c^* = -h[\sin^2\alpha \mathbf{c} \cdot (\nabla \times \mathbf{c}) + \sin\alpha\cos\alpha \mathbf{N}_0 \cdot (\nabla \times \mathbf{c}) + \cos^2\alpha e_k c_l \partial_k \partial_l u] \tag{11}$$

where $\mathbf{e} = \mathbf{N} \times \mathbf{c}$. The second $\mathbf{N}_0 \cdot (\nabla \times \mathbf{c})$ term in this expression integrates to the boundary if α is spatially

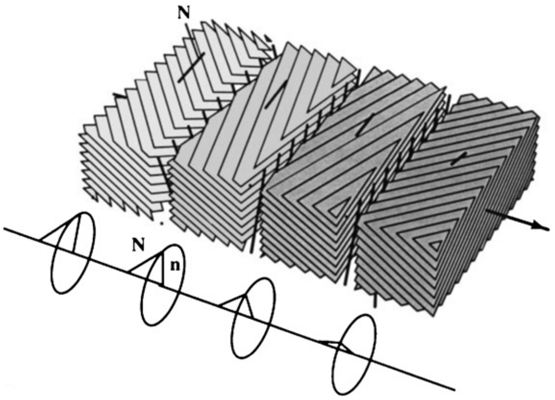


Figure 9
Bordeaux TGB_C phase. The layer normals **N** in smectic slabs rotate in discrete jumps across edge-screw grain boundaries in a cone with constant projection onto the pitch axis, which is perpendicular to grain boundary planes. The Frank director lies in plane perpendicular to the pitch axis and maintains a constant angle α with the layer normals.

uniform and there are no defects in the system, and it can often be neglected. It can, however, lead to the formation of spatially modulated stripe patterns in the smectic planes, especially in two-dimensional free-standing films (Selinger *et al.* 1993). The third term couples in-plane order to the curvature of smectic layers. It favors cylindrically bent layers with **c** at an angle of 45° to the cylindrical axis (Dahl and Lagerwall 1984) but it can be ignored in bulk smectics where layers are flat.

The first, $\mathbf{c}\cdot(\nabla\times\mathbf{c})$, term in Eqn. (11) is the most important in bulk systems. When **c** is expressed as $(\cos\phi, \sin\phi, 0)$, it can be written as

$$f^{*}_{\mathrm{ca}} = -H\partial_z\phi \tag{12}$$

where $H = h\sin^2\alpha$. This energy favors growth of θ along the z-direction. This growth is opposed by the SmC elastic energy $f^{\mathrm{el}}_{\mathrm{c}}$. The only term in $f^{\mathrm{el}}_{\mathrm{c}}$ that depends only on the z-derivatives of **c** is the term

$$f^{\mathrm{el}}_{\mathrm{c3}} = \frac{1}{2}A_3(\partial_z\mathbf{c})^2 = \frac{1}{2}A_3(\partial_z\phi)^2 \tag{13}$$

Thus, in the equilibrium state

$$\partial_z\phi = \frac{H}{A_3} \tag{14}$$

and the **c**-director precesses like the director in a cholesteric with its pitch axis along the layer normal.

Chirality destroys inversion and reflection symmetry and admits pseudoscalar terms, which change sign under spatial inversion, in the free energy. This has profound effects on the properties of chiral smectic-C liquid crystals. The cross product $\mathbf{e} = \mathbf{N}\times\mathbf{c}$ defines a vector perpendicular to **c** in the plane of the smectic layers. In achiral systems, **e** is equivalent to $-\mathbf{e}$. In chiral systems, **e** is not equivalent to $-\mathbf{e}$, and vectors such as the electric polarization will align with nonvanishing average value either along **e** or $-\mathbf{e}$. Thus, the lower symmetry induced by chirality will produce a nonvanishing dipole moment (Meyer *et al.* 1975) in the SmC* phase aligned perpendicular to both **c** and $\mathbf{n}_0$. This result can be obtained by observing that a term of the form

$$f^{*}_{p} = -h_p\mathbf{p}\cdot(\mathbf{N}\times\mathbf{c}) \tag{15}$$

is permitted in the free energy.

Bibliography

Abrikosov A A 1957 *Zh. Eksp. Teor. Fiz.* **32**, 1442 (*Soviet Physics JETP* **5**, 1174)

Dahl I, Lagerwall S T 1984 *Ferroelectrics* **84**, 215

Day A R, Lubensky T C, McKane A J 1983 *Phys. Rev. A* **27**, 1461

Goodby J, Waugh M A, Stein S M, Chin E, Pindak R, Patel J S 1988 *Nature* **337**, 449

Harris A B, Kamien R D, Lubensky T C 1999 *Rev. Mod. Phys.* **71**, 1745

Lagerwall S, Dahl I 1984 *Mol. Cryst. Liq. Cryst.* **114**, 151

Meyer R B, Liebert L, Strzelecki L, Keller P 1975 *J. Phys. (Paris) Lett.* **36**, L69

Navailles L, Barois P, Nguyen H 1993 *Phys. Rev. Lett.* **71**, 545; Navailles L, Barois P, Nguyen H 1994 *Phys. Rev. Lett.* **72**, 1300

Renn S, Lubensky T C 1988 *Phys. Rev. A* **38**, 2132

Renn S, Lubensky T C 1992 *Mol. Cryst. Liq. Cryst.* **209**, 349; Renn S 1992 *Phys. Rev. A* **45**, 953

Selinger J V, Wang Z -G, Bruinsma R F, Knobler C M 1993 *Phys. Rev. Lett.* **70**, 1139

Tinkham M 1996 *Introduction to Superconductivity*, McGraw-Hill, New York, 2nd edn

T. C. Lubensky
University of Pennsylvania, Philadelphia,
Pennsylvania, USA

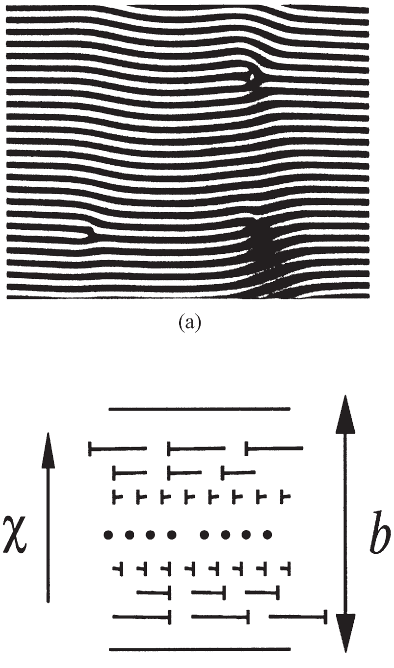
Cholesteric Liquid Crystals: Defects

For a long time the only defects to be seriously considered in condensed matter physics were dislocations in solids, mostly because they control the nonlinear (plastic) deformation properties of metals. But this limited point of view has been superseded by more general and profound considerations, which tie the nature of the defects in any material endowed with a nontrivial order parameter, (a solid, a liquid crystal, or even an isotropic phase like superfluid helium), to its various symmetries and/or to the topological properties of its order parameter (Michel 1980). One speaks of topologically stable defects (Toulouse and Kleman 1976). These extensions have been useful to researchers interested in cholesterics, which are rich in defects of different types. Although discovered as early as the beginning of the twentieth century (Friedel 1922), long before dislocations in solids, it is only since the early 1970s that defects in liquid crystals have been vigorously investigated.

1. The Classification of Defects Based on Symmetries

Dislocations are topological line defects: the translation symmetries are broken along a linelike region. Nematics display two types of topological defects, disclinations (line defects) and hedgehogs (point defects), both breaking the rotation symmetries of the phase. The cholesteric phase possesses line defects of two types: disclinations, as in nematics, and dislocations, as in solids. This classification, which rests on the cholesteric symmetries (Kleman and Friedel 1970), is a direct extension of the Volterra process for dislocations in solids (Friedel 1960). Topological considerations (Volovik and Mineyev 1977) show further that there are no point defects, contrary to nematics.

A cholesteric phase is characterized, at the length scale which is of interest here, by its pitch p, i.e., the distance over which the direction of the director **n** rotates by an angle of 2π about the cholesteric axis, denoted χ. A nematic phase is a cholesteric whose pitch is infinitely large. An easy way to obtain a pitch p large compared to the wavelength of light is to mix a nematic phase with a small proportion of Canada balsam, a chiral molecule. In these conditions the cholesteric structure is visible with the simple use of a polarizing microscope, Fig. 1(a). The direction of the molecule, or more precisely the local optical axis, will be noted **n**; χ is the cholesteric axis; the third direction is noted $\tau = \mathbf{n} \times \chi$. These three directions are directors, since changing any of them to its opposite does not change the cholesteric orientation. It follows that the repeat distance measured along χ is $p/2$, the half pitch. The variation of **n**, χ, and τ in space can be described by vector fields, which are called λ, χ, and τ fields, respectively.


(a)

(b)

Figure 1
(a) Cholesteric sample observed between crossed polarizers (8CB plus Canada balsam). Through each stripe the director **n** rotates by an angle of π about the cholesteric axis χ, which is constant and perpendicular to the stripe. Width of the stripe $p/2 \approx 2300$ nm (courtesy of Dr. C. Blanc). (b) Frank representation of the cholesteric orientation. The director **n** is represented as a nail whose head is supposed to be on the side of the reader and points to the back. The length of the nail is the projection on the plane of the sketch.

Defects in cholesterics are related to the symmetries of the structure, as follows: (i) due to the existence of three directors (χ, **n**, and τ), there are three types of disclinations, showing great similarities with disclinations in nematics, and (ii) the layered structure (of periodicity $p/2$) entails properties analogous to those of smectics: one expects to find dislocations, which break this periodicity.

This is indeed what is observed but the actual situation is somewhat more subtle, and shows interplays between the different types of defects. These interplays are better understood when described with the two languages, Volterra and topological. Besides, cholesteric defects, due to the 1D periodicity of the phase, have some close relation with focal conic domains, which are typical defects of lamellar phases.

2. Disclinations λ, τ, and χ

According to the orientation of the line defect with respect to the direction of the broken rotation, one distinguishes wedge disclinations, which are parallel to the broken rotation, twist disclinations, which are perpendicular to it, and mixed disclinations.

A defect which makes singular the two sets of directors, τ and χ, but leaves continuous the λ field, is called a 'λ disclination'. Such a situation is pictured in Fig. 2(a) for a wedge disclination of strength $k = -(1/2)$, featuring a breaking of symmetry of angle $2\pi|k| = \pi$ for the χ and the τ vector fields (both rotate by an angle of π when traversing a loop surrounding the disclination). Consequently there is some place inside the loop where the χ and the τ directors are not defined and where the defect line is located. This disclination is noted λ^-. The λ^+ line, of opposite strength $k = +(1/2)$, is pictured Fig. 2(b).

Figure 2(c) pictures a τ^-, i.e., a wedge line of strength $k = -(1/2)$, singular for the χ and λ fields and continuous for the τ field. A λ^- can be annihilated by collapse with a λ^+, Fig. 2(e). One can also feature χ lines, i.e., disclination lines which preserve the continuity of the χ-field; in fact, χ lines are directly related to dislocations and will be discussed apart.

It is possible to extend these results, which are reminiscent of lines in nematics, to the concept of escape in the third dimension, which is well understood in nematics. The answer is negative for a $k = 1$ λ^{2+}, Fig. 2(d), which cannot be made continuous for the two fields τ and χ altogether. The same result holds for a τ^{2+} or a χ^{2+}. On the other hand a λ^{4+}, a τ^{4+} and a χ^{4+} ($k = 2$) can be made entirely continuous. More generally, any line χ, λ, or τ, with strength $k = 2m$, m integer, can be made nonsingular by escape in the third dimension. This result can be fully justified in the frame of the topological theory of defects, see below.

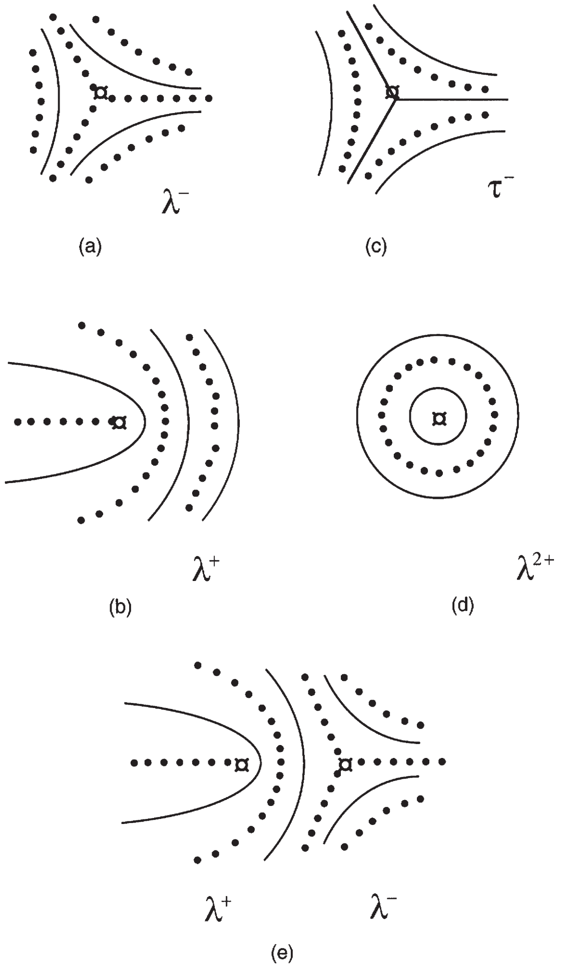


Figure 2
Wedge disclinations in a cholesteric: (a) λ^-, (b) $-\lambda^+$, (c) $-\tau^-$, (d) $-\lambda^{2+}$, and (e) $-\lambda^+$ and λ^- interacting (after Kleman and Lavrentovich 2000).

Figure 3(a) pictures a χ^+ wedge disclination (the χ field is continuous, and the λ and τ fields rotate by an angle of π about the line defect). The same object can as well be considered as a screw dislocation, since a π-rotation along the χ axis and a $p/2$ translation along the same axis add up to the same operation of symmetry. In the final configuration each cholesteric plane yields a 2D $k=1/2$, Fig. 3(a), whose configuration rotates helically along the χ line with a pitch p, the resulting Burgers vector is $b=-p/2$. Figure 3(b) extends to edge dislocations versus twist χ disclinations the equivalence just demonstrated between screw dislocations and wedge χ disclinations. Dislocations and χ-disclinations are fully equivalent, whatever the shape of the line. In the general case, one gets:

$$b=-kp \qquad (1)$$

3. *Effects of the Layer Structure*

3.1 *Dislocations*

An important property of dislocations in cholesterics is their possible splitting into λ or τ disclinations of opposite signs, at a distance multiple of $p/4$. Figure 3(c) pictures the same dislocation as in Fig. 3(b), whose core is now split into a λ^- and a τ^+ at a distance $p/4$ one from the other. The Burgers vector is $b=p/2$.

The splitting of a dislocation (equivalently: of a χ disclination) into two disclinations λ or τ of strength $|k|=1/2$, i.e., of rotation $\pm\pi$, relates to the fact that the product of two opposite π rotations along two parallel axes is a translation (Kleman and Friedel 1970). Split edge dislocations have been observed and studied in Grandjean-Cano wedge samples. The singularities (observed end-on) of Fig. 1(a) are edge dislocations.

3.2 *Focal Conic Domains, Polygonal Textures, and Oily Streaks*

Most frequently, cholesterics present domains analogous to the focal domains of smectics, as one might expect from the existence of a 1D periodicity. However the layers can suffer large thickness distortions (contrarily to smectics), since it is possible to let the pitch vary in a large range at the moderate expense of some twist energy K_2. As a rule, the cholesteric layers are saddle-shaped in focal conic domains. Polygonal textures are sets of domains where the layers, observed on the side, are practically parallel, separated by boundaries marked by line defects (Fig. 4). Oily streaks are complex aggregates of edge dislocations of large, opposite, Burgers vectors, present in samples whose boundaries are parallel to the cholesteric layers (Fig. 5). They partition the sample in ideal domains of flat layers.

3.3 *Robinson Spherulites*

Semiflexible cholesteric biopolymers (but also some thermotropic cholesterics, Fig. 6), have other types of layered textures, the so-called Robinson spherulites (Pryce and Frank 1958). The layers are approximately along concentric spheres. The molecular orientation is singular, since it is impossible to outline a continuous field of directors on a sphere. The total strength of these singularities is $k=2$. And indeed observations tell us that either a $k=1$ singularity goes through the whole spherulite along a diameter, or a $k=2$ line defect extends from the surface to some point inside (generally the center of the spherulite). The $k=2$ line is believed to relax to a nonsingular configuration, by escape (Bouligand and Livolant 1994).

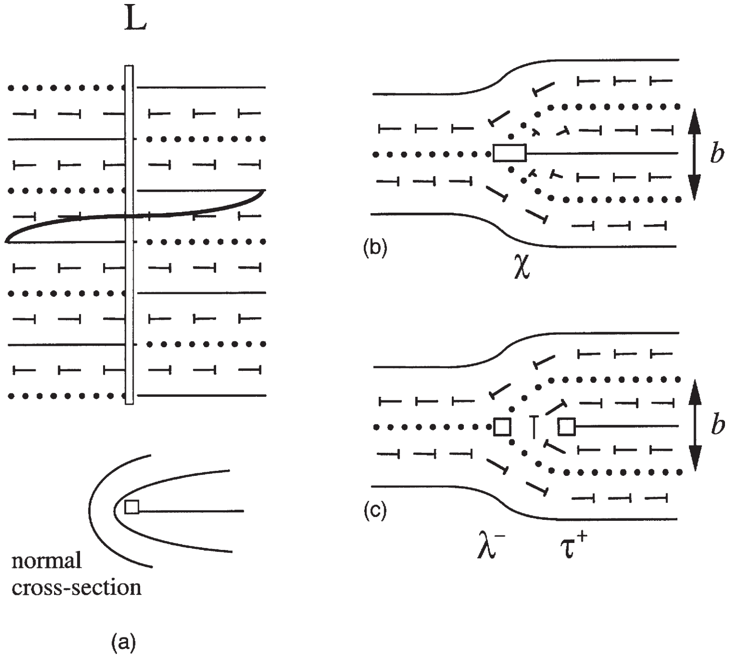


Figure 3
Equivalences: (a) χ^+ wedge disclination≡screw dislocation, χ twist disclination≡edge dislocation (after Kleman and Lavrentovich 2000).

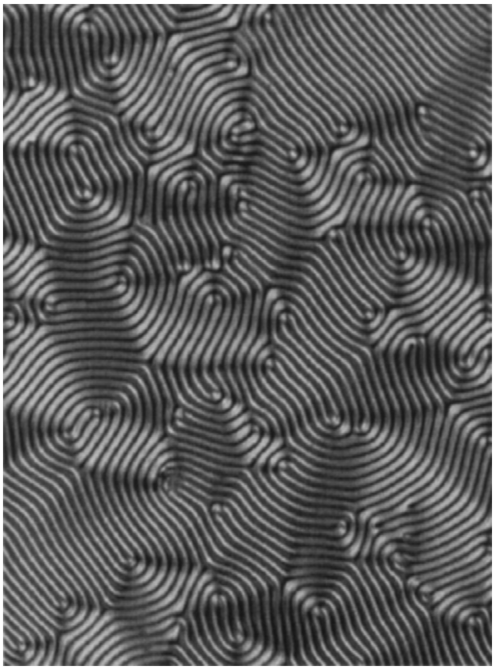


Figure 4
Polygonal texture in 8CB plus Canada balsam, crossed polarizers, $p/2 \approx 2000\,\mathrm{nm}$ (courtesy Dr. C. Blanc).

4. *Double Twist and Frustration*

Among the elastic coefficients which enter in the free energy of the cholesteric phase, the K_{24} term, which is generally discarded because it amounts to a surface term, plays a very special role. To illustrate this point, consider a string with helical strands, the central strand being straight and the successive strands making with the central strand an angle $\psi(R)$ which increases with R, $\psi(0) = 0$. This string is an adequate representation of the geometry of double-twist. The cholesteric molecules (the director **n**) are aligned along the strands. Note that the strands are equidistant, so that this geometry describes also a helical dense packing of chiral molecules (Kleman 1985). The double-twisted cylinder cannot extend to a radius large compared to the natural pitch: it gets *frustrated.* Indeed, let $\psi(R)$ be the angle of **n** with the axis of the string. The free energy density can be written:

$$f = \tfrac{1}{2}K_2\left(q - \frac{\partial\psi}{\partial r} - \frac{1}{r}\sin\psi\,\cos\psi\right)^2 + \tfrac{1}{2}K_3\frac{\sin^4\psi}{r^2} - \frac{K_{24}}{r}\frac{\mathrm{d}}{\mathrm{d}r}(\sin^2\psi) \qquad (2)$$

Integrating f, it is apparent that the K_{24} term contributes to the energy of a cylinder of matter of radius R by the quantity $F_{24} = -2\pi K_{24}\sin^2\psi(R)$ which is

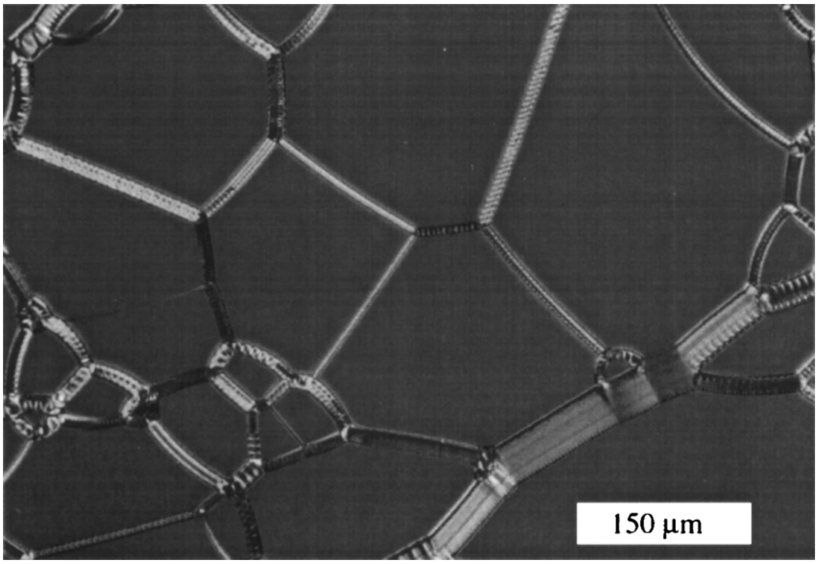
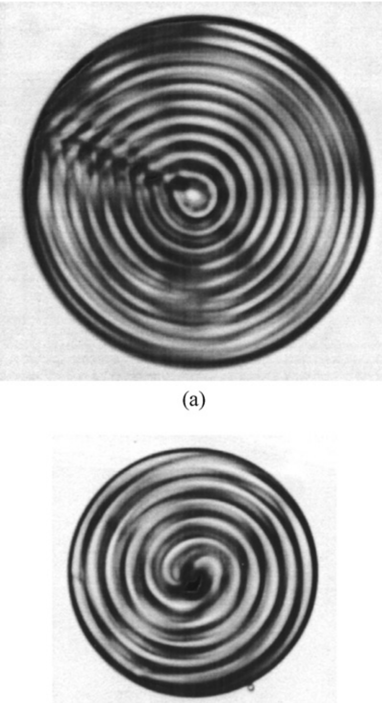


Figure 5
Oily streaks in a small pitch cholesteric, crossed polarizers, $p/2 \approx 3000$ nm (courtesy of Professor O. D. Lavrentovich).


(a)

(b)

Figure 6
Robinson spherulites of 8CB plus Canada balsam, polarizers slightly uncrossed. Disclination line $k=2$. (a) In the plane of observation and (b) perpendicular to the plane of observation (courtesy of Dr. C. Blanc).

negative for any value of $\psi(R) \neq \pi m$, when K_{24} is positive, and maximum in modulus for $\psi = \pi/2$. Hence, the *nucleation* of a finite double-twisted geometry is favored, in particular when K_1 is large compared to K_3. The $k=2$ singularity-free configuration in a cholesteric is a region of concentration of double-twist. If K_{24} is positive and very large, the cylindrical geometry can become stable versus the cholesteric phase: this is the origin of the blue phases in which finite double-twisted cylinders arrange in a 3D structure, separated by defects.

Another double-twisted (and consequently quasi dense packed) geometry is met in the DNA arrangement of the chromosome of a microscopic alga, *Prorocentrum Micans*, whose cholesteric structure is well documented (Livolant and Bouligand 1978). The double-twisted region is bound by two $k=1/2$ disclination lines which rotate helically about the chromosome axis (Kleman 1987), and so limit the size of the chromosome. While the nucleation of the $k=2$ geometry is made easier with $K_{24}>0$, the opposite condition holds with the chromosome.

5. *The Classification of Defects Based on Topology*

The topological classification of defects in cholesterics rests on the topology of the order parameter, whose geometry is that of the orthogonal set $\{\mathbf{n}, \chi, \tau\}$. The rotations of this trihedron generate the group $SO(3)$, which is also the 3D sphere S_3 with antipodal points identified. The translations, which are equivalent to χ rotations, can be discarded. A loop surrounding a

Table 1
Correspondence between the Volterra (λ, τ, and χ) classification and the topological classification of defect lines in cholesteric liquid crystals (m is an integer).

C_0	$\bar{C}_0$	$C_\lambda(=C_x)$	$C_\tau(=C_y)$	$C_\chi(=C_z)$
$\lambda(2m)$	$\lambda(2m+1)$	$\lambda(m+(\frac{1}{2}))$		
$\tau(2m)$	$\tau(2m+1)$		$\tau(m+(\frac{1}{2}))$	
$\chi(2m)\equiv\{b=-2mp\}$	$\chi(2m+1)\equiv\{b=-(2m=1)p\}$			$\chi(m=\frac{1}{2})\equiv\{b=-(m+(\frac{1}{2}))p\}$

2π rotation defect ($k=\pm 1$) maps on S_3 along a path which joins two antipodal points. Such a path—a loop in $SO(3)$—being nontrivial in the sense that it cannot be shrunk smoothly on $SO(3)$ to a point, a $k=\pm 1$ defect necessarily owns a singular core. On the other hand, a loop in $SO(3)$ representing a 4π rotation defect ($k=\pm 2$) can be shrunk to a point on $SO(3)$, from which we infer that a 4π disclination requires no singular core. This important result is also true for ^{3}He and biaxial nematics (Anderson and Toulouse 1977) whose order parameter has the same geometry. The whole group of defects is isomorphic to the first homotopy group of the order parameter space R, which is $SO(3)$ factorized by the four-element point group D_2 of π-rotations about the directions $\mathbf{n}$, χ, and τ:

$$R = SO(3)/D_2 \tag{3}$$

The fundamental group $\pi_1(R)=\mathrm{Q}$ is the eight-element quaternions group, whose elements fall into five conjugacy classes $C_0=\{I\}$, $\bar{C}_0=\{J\}$, $C_x=\{r,-r\}$, $C_y=\{s,-s\}$ and $C_z=\{t,-t\}$, each class corresponding to an irreducible Volterra defect (see Table 1). The nonexistence of point defects follows from the fact that the second homotopy group is trivial: $\pi_2(R)=0$. The noncommutativity of Q yields interesting entanglements effects between defects (Poenaru and Toulouse 1977). They can also be discussed, although less systematically, via the Volterra process (Kleman and Lavrentovich 2000).

Since the $\mathbf{n}$ director field alone has a material reality, it is tempting to consider its topology independently of the two other directors. This approach has led to the notion of distortions with double topological character and to a classification of topologically stable configurations (without singularities) based on the Hopf mapping (Bouligand *et al.* 1978).

6. Concluding Remarks

The concept of chirality has taken a great extension in soft matter physics, due to the manifold presence of chiral molecules. Defects and textures in cholesterics, apart from their fundamental physical interest, are also an active field of research in cholesteric phases of biopolymers, like DNA, xanthan, etc., (Bouligand 1998). Chirality appears under different ordered realizations (e.g., in the many twist grain boundaries phases) in which the notion of defect plays a crucial role. One can therefore expect that investigations on chirality cannot but increase in the future. And since cholesterics are in many respects the simplest chiral ordered systems, there is all the more reason to pursue the effort to fully understand them.

Bibliography

Anderson P W, Toulouse G 1977 Phase slippage without vortex cores: vortex textures in superfluid ^{3}He. *Phys. Rev. Lett.* **38**, 508–11

Bouligand Y 1998 Defects and textures. In: Demus D, Goodby J, Gray G W, Spiess H -W, Vill V (eds.) *Handbook of Liquid Crystals, Fundamentals, Vol. 1.* Wiley-VCH, Weinheim, Germany, pp. 406–53

Bouligand Y, Derrida B, Poenaru V, Pomeau Y, Toulouse G 1978 Distortions with double topological character: the case of cholesterics. *J. Phys.* **39**, 863–7

Bouligand Y, Livolant F 1994 The organisation of cholesteric spherulites. *J. Phys.* **45**, 1899–923

Chandrasekhar S 1992 *Liquid Crystals.* Cambridge University Press, Cambridge

Friedel G 1922 Les états mésomorphes de la matière. *Ann. Phys. (Paris)* **9**, 273–374

Friedel J 1960 *Dislocations.* Pergamon, Oxford

Kleman M 1985 Frustration in polymers. *J. Phys. Lett. (Paris)* **46**, 723–32

Kleman M 1987 Effects of frustration in liquid crystals and polymers. *Phys. Scr. T* **19**, 565–72

Kleman M, Friedel J 1970 Application of dislocation theory to liquid crystals. In: *Fundamental Aspects of the Theory of Dislocations.* NBS Special Publications 317, pp. 607–36

Kleman M, Lavrentovich O D 2000 *Introduction to the Physics of Soft Matter.* Springer, New York

Livolant F, Bouligand Y 1978 New observations on the twisted arrangement of Dinoflagellate chromosomes. *Chromosoma* **68**, 21–44

Michel L 1980 Symmetry defects and broken symmetry. Configurations. Hidden symmetry. *Rev. Mod. Phys.* **52**, 617–51

Poenaru V, Toulouse G 1977 The crossing of defects in ordered media and the topology of manifolds. *J. Phys.* **38**, 887–95

Pryce M H L, Frank F C 1958 Liquid crystalline structures in solutions of a polypeptide, part 2. *Disc. Faraday Soc.* **25**, 29–42

Toulouse G, Kleman M 1976 Principles of classification of topologically stable defects in ordered media. *J. Phys. Lett. (Paris)* **37**, 149–51

Volovik G E, Mineyev V P 1977 Investigation of singularities in superfluid He3 and liquid crystals by the homotopic topology methods. *Sov. Phys. JETP Lett.* **45**, 1186–96

M. Kleman
Université Pierre-et-Marie-Curie, Paris, France

Clay-based Polymer Nanocomposites

A clay comprises silicate layers having a 1 nm thick planar structure. It has been shown that the silicate layers can be dispersed at the molecular level (nanometer level) in a polymer matrix with the polymer existing between the silicate layers (two-dimensional space). Polymer and clay nanocomposite materials produced in this manner are called polymer–clay hybrid materials. In this article some materials of this type are described, in categories based on synthetic methods. Five methods of synthesis of clay-based polymer nanocomposites will be described on the basis of the relationship between an ammonium cation-exchanged clay and monomer (or polymer). The methods are: (i) monomer intercalation, (ii) monomer modification, (iii) co-vulcanization, (iv) common solvent, and (v) polymer intercalation.

1. Classification of Producing Polymer–Clay Hybrid Material According to the Synthetic Method Employed

1.1 Monomer Intercalation Method

A polymerization producing nylon 6 is the ring-opening polymerization of ε-caprolactam. It can occur in the presence of clay after ε-caprolactam is intercalated into clay galleries, the silicate layers being dispersed uniformly in the nylon 6 matrix. We found that organophilic clay which had been ion-exchanged with 12-aminododecanoic acid was swollen by molten ε-caprolactam (basal spacing being expanded from 1.7 nm to 3.5 nm) (Usuki *et al.* 1993a). In the clay galleries the ε-caprolactam was polymerized, and the silicate layers were dispersed in the nylon 6 to give a nylon 6 clay hybrid (NCH) (Usuki *et al.* 1993b). This is the first example of an industrial clay-based polymer nanocomposite. Figure 1 shows a schematic representation of the polymerization.

The modulus of NCH increased to 1.5 times that of nylon 6, the heat distortion temperature increased to 140 °C from 65 °C, and the gas barrier effect was doubled at a low loading (2 wt.%) of clay (Kojima *et al.* 1993b).

There is another example, in which ε-caprolacton is polymerized in the clay galleries in the same manner. In this case, the gas permeability decreased to about 20% under 4.8 vol.% (12 wt.%) clay addition (Messersmith and Giannelis 1995). There is yet another example of an epoxy resin clay nanocomposite. In this case, the tensile strength and modulus increased drastically on 2–20 wt.% clay addition (Lan and Pinnavaia 1994).

1.2 Monomer Modification Method

In one example, a quaternary ammonium salt of dimethylaminoacrylamide (Q) was ion-bonded to silicate layers, and ethyl acrylate (EA) and acrylic acid (Aa) were copolymerized in the clay galleries. The ratio between EA and Aa was 10:1 (mole ratio). Each of the four kinds of acrylic resin–clay hybrid were polymerized. The clay content was 1, 3, 5, and 8 wt.% on the basis of solid acrylic resin. A suspension above 3 wt.% clay addition acted as pseudoplastic fluid (Tables 1 and 2 give details of compositions). The acrylic resin clay hybrid films cross-linked by

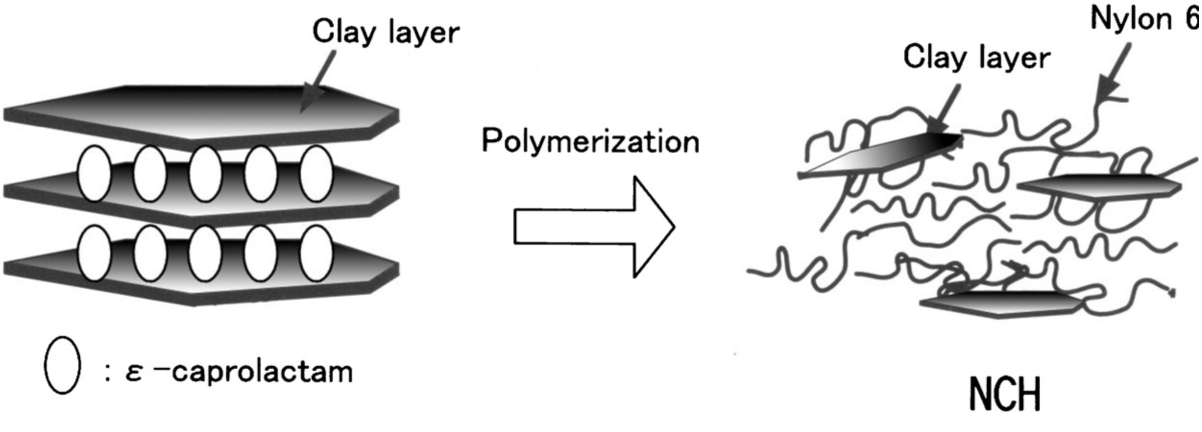


Figure 1
Schematic diagram of polymerization to NCH.

Table 1
Composition of acrylic resin.[a]

Acrylic resin	Ethyl acrylate, EA (g)	Acrylic acid, Aa (g)	Dimethylamino acrylamide, Q (g)	Butyl cellosolve (g)	*NV, calculated*[b] (%)
A	144.7 (90.8)	10.4 (9.08)	0.51 (0.12)	157.7	46.0
B	144.7 (90.6)	10.4 (9.04)	1.60 (0.36)	157.7	46.1
C	144.7 (90.4)	10.4 (9.04)	2.54 (0.56)	157.7	46.2
D	144.7 (90.1)	10.4 (9.01)	4.13 (0.89)	157.7	45.0

[a] Figures in parentheses are mole ratios. [b] NV: nonvolatile concentration.

Table 2
Composition of acrylic resin–clay hybrid.

Acrylic resin–clay hybrid	Acrylic resin (g)	NV, measured[a] (%)	Deionized water (g)	Montmorillonite–water suspension, 4 wt.% (g)	*NV* in montmorillonite[a] (%)
A	a 14.0	44.9	16.6	1.6	1.0
B	b 14.0	45.2	11.6	4.9	3.1
C	c 13.4	45.3	10.5	7.5	4.9
D	d 14.5	43.9	6.4	12.7	8.0

[a] NV: nonvolatile concentration.

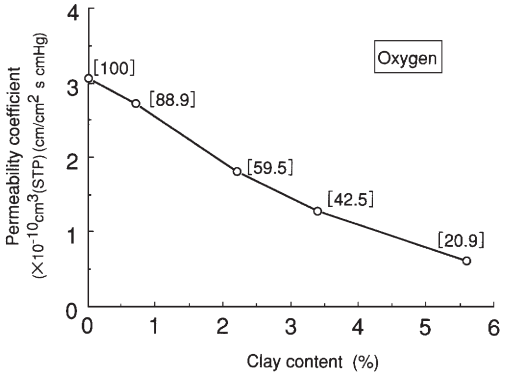


Figure 2
Permeability coefficient of acrylic resin clay hybrid film.

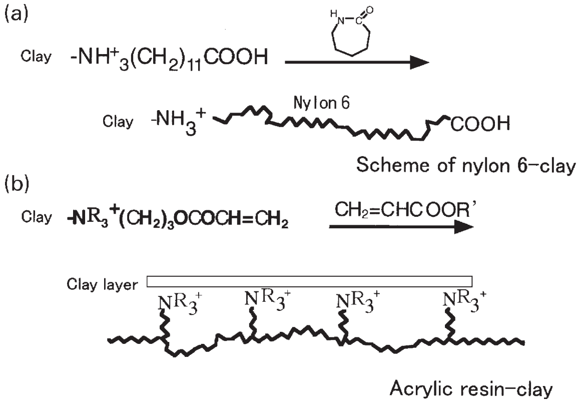


Figure 3
Scheme of NCH and acrylic resin clay hybrid.

melamine were transparent, and the gas permeability of their film (Fig. 2) decreased to about 50% on 3 wt.% clay addition (Usuki *et al.* 1995).

There are also some other reports of acrylic resin–clay hybrids. A poly(methyl methacrylate) clay hybrid was synthesized using a modified organophilic clay in the same manner (Biasci *et al.* 1994), and by emulsion polymerization (Choo and Jang 1996). Figure 3 shows a schematic representation of this polymerization method.

1.3 Co-vulcanization

The basal spacing of organophilic clay ion-bonding nitrile rubber (NBR) oligomer with telechelic amino groups was expanded 0.5 nm from its initial spacing (1.0 nm) (Moet *et al.* 1994). After this, high molecular weight NBR was kneaded with this organophilic clay and vulcanized with sulfur. It produced an NBR clay hybrid consisting of dispersed clay and co-vulcanized high molecular weight NBR and NBR oligomer (Fukumori *et al.* 1991). Its permeability to hydrogen and water decreased to 70% on adding 3.9 vol.% clay (Kojima *et al.* 1993a). Figure 4 shows a schematic representation of this production method.

1.4 Common Solvent Method

In the case of synthesis of polyimides, the polymerization solvent for the polyamic acid precursor of polyimide is usually dimethyl acetamide (DMAC). We found that a clay ion-exchanged dodecyl ammonium ion could be dispersed in DMAC homogeneously. The solution of this organophilic clay and DMAC was added to a DMAC solution of polyamic acid. The film was cast from a homogeneous mixture of clay and polyamic acid, and was heated at 300 °C to get the desired polyimide clay hybrid film. The permeability of water decreased to 50% on adding 2.0 wt.% clay (Yano *et al.* 1993). It was confirmed that the permeability of carbon dioxide also decreased by half (Lan *et al.* 1994). Figure 5 shows a schematic representation of this method.

1.5 Polymer Intercalation Method

Polypropylene (PP) clay hybrids could not be synthesized easily because PP is hydrophobic and has poor miscibility with clay silicates. Dioctadecyldimethyl ammonium ion was used as a modifier for the clay and a polyolefin oligomer was used so that the clay became more compatible. Organophilic clay, polyolefin oligomer, and PP were blended using an extruder at 200 °C. It was confirmed by transmission electron microscopy that the clay was dispersed in a monolayer state in the PP matrix. Thus PP was directly intercalated into the clay gallery (Usuki *et al.* 1997). Figure 6 shows a schematic representation of this polymerization.

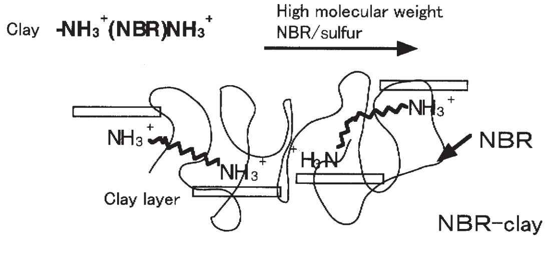


Figure 4
Scheme of NBR clay hybrid.

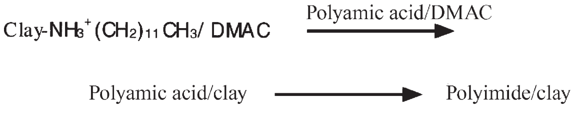


Figure 5
Scheme of polyimide clay hybrid.

There is also a direct-intercalation process in which PP is modified using maleic anhydride, followed by melt compounding (Kawasumi *et al.* 1997). It is a useful process from an industrial standpoint. There are also some studies showing that polymer is intercalated directly into the clay galleries. There is a report that intercalation of nylon into clay galleries was successful (Takemura *et al.* 1996). The clay silicates may not have been dispersed uniformly, judging by the physical properties exhibited.

2. *Functional Clay Hybrids*

The examples described above involved organic polymer clay hybrids which are aimed at high performance structural polymer materials with high strength, high modulus, and high gas barrier properties. This hybrid technique should also be applicable to various other types of materials, including low molar mass compounds. This section presents one of the unique attempts to obtain novel hybrid materials based on a low molar mass liquid crystal and organized clay mineral (liquid crystal clay composite, LCC). Clay exchanged with 4-cyano-(4′-biphenyloxy)undecyl ammonium ion was used to enhance the miscibility between clay and the liquid crystal. The LCC (clay content 1.25 wt.%) cell displayed a reversible and bistable electro-optical effect based on light scattering which could be controlled by changing the frequency of an electric field (Kawasumi *et al.* 1996). Figure 7 shows a typical change in the light transmittance of the LCC (clay content 1.25 wt.%) when the cells were subjected to electric fields.

This new material would be a potential candidate for advanced applications such as a light controlling glasses, high information display devices which do not require active addressing devices, erasable optical storage devices, and so on.

3. *Possible Future Subjects*

The above examples illustrate the synthesis methods and the properties of clay-based nanocomposites including polymers and liquid crystals. A modification

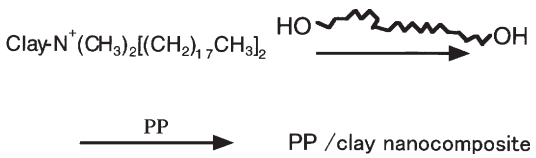


Figure 6
Scheme of PP clay hybrid.

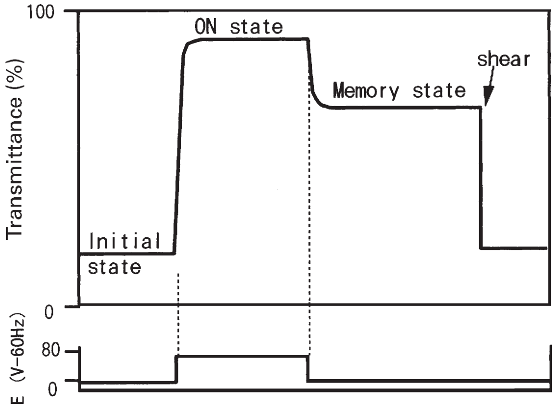


Figure 7
Light transmittance of LCC.

of the clay surface plays a principal part in the technology. Other modifications of monomer and/or polymer, or an addition of adequate compatibilizer for organic materials and clay could open a new methodology of clay nanocomposite production. Future subjects are considered to be:

(i) Clay-based nanocomposite technology involving polymer alloys.

(ii) Investigations of clay-based nanocomposites involving new properties such as transparency, recyclability, fire retardation (Gilman and Kashiwagi 1997), and biodegradation.

(iii) Identification of the most suitable clay manufacturing technique for each polymer.

Bibliography

Biasci L, Aglietto M, Ruggeri G, Ciardelli F 1994 Functionalization of montmorillonite by methyl methacrylate polymers containing side-chain ammonium cations. *Polymer* **35**, 3296–304

Choo D, Jang L W 1996 Preparation and characterization of PMMA–clay hybrid composite by emulsion polymerization. *J. Appl. Polym. Sci.* **61**, 1117–22

Fukumori K, Usuki A, Sato N, Okada A, Kurauchi T 1991 Rubber–clay molecular composites: nitrile rubber/layered silicate system. In: *Proc. 2nd Japan Int. SAMPE Symp.* SAMPE, Covina, CA, pp. 89–96

Gilman J W, Kashiwagi T 1997 Nanocomposites: a revolutionary new flame retardant approach. *ADDITIVES* '97, Executive Conference Management

Kawasumi M, Hasegawa N, Kato M, Usuki A, Okada A 1997 Preparation and mechanical properties of polypropylene–clay hybrids. *Macromolecules* **30**, 6333–8

Kawasumi M, Usuki A, Okada A, Kurauchi T 1996 Liquid crystalline composite based on a clay mineral. *Mol. Cryst. Liq. Cryst.* **281**, 91–103

Kojima Y, Fukumori K, Usuki A, Okada A, Kurauchi T 1993a Gas permeabilities in rubber–clay hybrid. *J. Mater. Sci. Lett.* **12**, 889–90

Kojima Y, Usuki A, Kawasumi M, Okada A, Fukushima Y, Kurauchi T, Kamigaito O 1993b Mechanical properties of nylon 6–clay hybrid. *J. Mater. Res.* **8**, 1185–9

Lan T, Kaviratna P D, Pinnavaia T J 1994 On the nature of polyimide–clay hybrid composites. *Chem. Mater.* **6**, 573–5

Lan T, Pinnavaia T J 1994 Clay-reinforced epoxy nanocomposites. *Chem. Mater.* **6**, 2216–9

Messersmith P B, Giannelis E P 1995 Synthesis and barrier properties of poly(ε-caprolacton)–layered silicate nanocomposites. *J. Polym. Sci., Polym. Chem.* **33**, 1047–57

Moet A, Akelah A, Salahuddin N, Hiltner A, Baer E 1994 Layered silicate/ATBN nanocomposites. In: Gonsalves K E, Chow G, Xiao T D, Cammarata R C (eds.) *Mat. Res. Soc. Symp. Proc.* Materials Research Society, pp. 163–70

Takemura K, Hosokawa T, Tamura K, Iniue H 1996 Polymer–clay composites. *Japan–China Seminar on Advanced Engineering Plastics, Polymers, Alloys and Composites*, Society of Polymer Science, pp. 32–5

Usuki A, Kato M, Okada A, Kurauchi T 1997 Synthesis of polypropylene–clay hybrid. *J. Appl. Polym. Sci.* **63**, 137–9

Usuki A, Kawasumi M, Kojima Y, Okada A, Kurauchi T, Kamigaito O 1993a Swelling behavior of montmorillonite cation-exchanged for ω-amino acids by ε-caprolactam. *J. Mater. Res.* **8**, 1174–8

Usuki A, Kojima Y, Kawasumi M, Okada A, Fukushima Y, Kurauchi T, Kamigaito O 1993b Synthesis of nylon 6–clay hybrid. *J. Mater. Res.* **8**, 1179–84

Usuki A, Okamoto K, Okada A, Kurauchi T 1995 Synthesis and properties of acrylic resin–clay hybrid. *Kobunsi Ronbunshu* **52**, 728–33

Yano K, Usuki A, Okada A, Kurauchi T, Kamigaito O 1993 Synthesis and properties of polyimide–clay hybrid. *J. Polym. Sci., Polym. Chem.* **31**, 2493–8

A. Usuki
Toyota Central R&D Labs, Aichi, Japan

Cobalt Alloys: Alloying and Thermomechanical Processing

Cobalt alloys have been in use since 1907 when Elwood Haynes obtained the first patents on cobalt–chromium compositions. This article describes the nature of cobalt alloys and provides the basis for the selection of elements that are used in the alloys of commercial importance. The article also provides a brief description of the manufacturing methods that are employed.

1. Alloying

Like its neighbors in the periodic table (nickel and iron), cobalt can be alloyed with significant quantities of elements such as chromium, molybdenum, and tungsten to create very useful engineering materials. Indeed, cobalt imparts to its alloys some extremely important properties, such as resistance to several forms of wear, creep and fatigue strength at high temperatures, and resistance to sulfidation.

The resistance to wear of cobalt alloys derives partly from the fact that they undergo a structural transformation during mechanical deformation. Cobalt alloys also benefit from mechanical twinning and planar slip propensities.

Cobalt itself exhibits a h.c.p. structure at temperatures up to 417 °C. At higher temperatures (up to the melting point) the atomic structure is f.c.c. Alloying elements such as chromium, molybdenum, and tungsten increase the transformation temperature, i.e., they stabilize the h.c.p. structure, whereas elements such as nickel and iron suppress the transformation temperature and thus stabilize the f.c.c. form.

Importantly, the transformation of cobalt and its alloys from f.c.c. to h.c.p. upon cooling from the molten state or some high temperature in excess of the transformation point is extremely sluggish, and, in most cases, cobalt and its alloys exist in a metastable f.c.c. form at room temperature, unless they have been subjected to mechanical deformation. It is possible to generate the h.c.p. phase in alloys with sufficiently high transformation temperatures by holding the alloys close to this temperature, but this requires considerable time. As already inferred, the easiest way to generate the h.c.p. structure (at least partially) is by cold working. The formation of so-called "h.c.p. platelets" is believed to progress through stacking fault coalescence. The stacking fault energy of the metastable f.c.c. cobalt alloys is exceptionally low, therefore wide stacking faults (which are atomically thin h.c.p. plates) are abundant in cold worked materials.

There are several cobalt alloy types. The first cobalt alloys were designed primarily for wear resistance and contained high carbon contents to encourage the formation of carbides in the microstructure during alloy solidification. Since these carbides tend to limit hot ductility, hence restrict hot forging and hot rolling, the high-carbon alloys are generally used in cast or weld overlay form, rather than in the form of wrought products (sheets, plates, bars, etc.).

Next to be developed were the cast cobalt alloys for dental and biomedical use, i.e., to resist wear and corrosion from body fluids. These alloys contained less carbon than the original materials. Experimentation with these alloys under high-temperature conditions led to the evolution of the cast and wrought high-temperature alloys, which provide high-temperature strength, oxidation resistance, and sulfidation resistance.

A few cobalt alloys have been designed with general resistance to aqueous corrosion in mind. These alloys have been optimized for corrosion resistance, and take advantage of the inherent wear resistance and/or the strengths that can be attained by cold working.

Cobalt alloys, therefore, may be grouped according to their primary uses, as follows:

(i) Wear
(ii) High temperature
(iii) Biomedical
(iv) Aqueous corrosion.

It is difficult to generalize about product forms, however, since some groups involve several forms (wrought, cast, weld overlay, and powder metallurgy). Also, even if optimized for a particular end use, all cobalt alloys share the characteristics of good wear resistance, good high-temperature strength, and reasonable resistance to aggressive environments (liquid and gaseous).

Although many of the same alloying elements are used to optimize cobalt alloys for their different end uses, their roles are not necessarily the same. Chromium, for example, is used to impart resistance to oxidation and sulfidation to the high-temperature alloys and passivity to the aqueous corrosion alloys, whereas its primary function in the wear alloys is as a carbide former. The chromium that remains in solution provides environmental resistance, some solid solution strengthening, and increases the f.c.c. to h.c.p. transformation temperature, thus having an impact on the transformation tendencies under the action of mechanical deformation.

The carbon contents of the wear alloys vary from about 0.5 wt.% to 3.3 wt.%. This much carbon results in significant carbide weight percentages in the microstructures; for example, approximately 13 wt.% carbide (predominantly chromium-rich M_7C_3) at 1.1 wt.% carbon, and 29 wt.% carbide (predominantly chromium-rich M_7C_3 and tungsten-rich M_6C) at 2.5 wt.% carbon. As will be discussed, these microstructural carbides are of great benefit under low stress abrasion conditions.

In addition to chromium, the wear alloys typically contain tungsten. The primary function of this element is to provide solid solution strengthening and to promote transformation tendencies (being a stabilizer of the h.c.p. form); it also partitions to the carbides in the higher carbon alloys.

A small, but commercially significant, subset of cobalt wear alloys is alloyed with chromium, molybdenum, and silicon, rather than chromium, tungsten, and carbon. Instead of relying on carbides to optimize low stress abrasion resistance, these alloys rely on the formation of Laves phase (an intermetallic compound, promoted by the presence of silicon). In these alloys, the primary function of molybdenum is Laves phase formation. Unfortunately, their ductility is limited to such an extent that large castings and weld overlays are impractical. Indeed, these alloys have been most successful in powder form, as consumables for plasma spraying.

With regard to high-temperature alloys, it has already been stated that the primary function of chromium is to provide resistance to the environment. In oxidizing, high-temperature environments, protection is afforded by Cr_2O_3 films which form on the surfaces. In high-temperature, sulfur-bearing environments in which the oxygen partial pressure is not very low, Cr_2O_3 films are again important, since they provide protection in the early stages of the process. Chromium is also an effective alloying addition because its sulfides are more stable than those of cobalt and nickel and they have much higher melting points. Cobalt alloys are intrinsically more resistant to sulfidation than nickel alloys, possibly because of the lower diffusivity of sulfur and the higher melting point of cobalt sulfide. In corrosive, aqueous media, chromium is beneficial in that it forms protective (oxide or hydroxide) films, providing oxidizing species are present.

Tungsten is used to strengthen high-temperature cobalt alloys. It is a very effective solid solution strengthener because of its large atomic size as well as its tendency to lower the stacking fault energy, thereby making dislocation cross-slip more difficult. Molybdenum is also an effective strengthener for the same reasons, although it is used mostly in the aqueous corrosion alloys to enhance their resistance to reducing acids such as hydrochloric and dilute sulfuric.

The carbon contents of the high-temperature and aqueous corrosion alloys are much lower than those of the cobalt wear alloys. Most of the high-temperature alloys, for example, have carbon contents of about 0.1 wt.%, which produces a sparse dispersion of primary carbides, which effectively pin the grain boundaries during high-temperature exposure. In most alloys secondary $M_{23}C_6$ carbides precipitate on glide dislocations during high-temperature creep, thereby enhancing resistance to further deformation. The aqueous corrosion alloys have even lower carbon contents, within the soluble range, to minimize the formation of carbides which can reduce aqueous corrosion resistance, especially when they precipitate in the grain boundaries of heat-affected zones during welding.

Apart from chromium, tungsten, and molybdenum, the most important elements in the high-temperature and aqueous corrosion alloys are nickel and iron. As previously mentioned, these are f.c.c. stabilizers and are used to counteract the effects of chromium, tungsten, and molybdenum. In other words, they are used to stabilize the f.c.c. structure and indirectly to make the alloys more amenable to processing, particularly at room temperature (by reducing the transformation tendencies under the action of mechanical deformation). The only problem with nickel and iron is that, in reducing the transformation tendencies, they reduce resistance to wear. For the high-temperature materials this is not an important issue, but for the aqueous corrosion materials (which are typically only used when a combination of corrosion resistance and wear resistance is needed) it is important. In the aqueous corrosion alloys, therefore, the h.c.p. and f.c.c. stabilizers are carefully balanced, and other elements, such as nitrogen, are used to enhance properties.

In the past, attempts were made to develop age-hardenable cobalt-based alloys for high-temperature applications using additions of titanium or tantalum to form the compounds Co_3Ti and Co_3Ta, analogous to $Ni_3(Al,Ti)$ in nickel-based alloys. However, due to problems with the thermal stability of the precipitates because of their higher lattice mismatch and their much lower solvus temperatures, they were easily surpassed by the nickel-based superalloys, and none of these alloys were commercially successful. In cast alloys, additions of elements such as tantalum, niobium, and zirconium are made for solid solution strengthening and to form MC-type carbides which enhance high-temperature strength. In both cast and wrought alloys, small additions of boron are also used to promote creep and stress-rupture strength.

As mentioned previously, chromium is the principal alloying element used to provide oxidation resistance. Attempts have been made to use aluminum additions in order to promote the formation of protective Al_2O_3 surface films, but such additions result in the formation of β-CoAl which significantly reduces ductility. Minor additions ($<1\%$) of elements such as manganese and silicon are used to promote the formation of very protective spinel oxides. Small additions of rare-earth elements such as lanthanum have also proved to be very effective in reducing oxide scale spallation.

2. *Processing*

The melting and processing of wrought cobalt alloys is similar to that of wrought nickel alloys. Typically, they are melted in arc furnaces, refined in argon–oxygen decarburization vessels, poured into large electrode molds, electroslag remelted, then hot forged and hot rolled to plates and bars. Wrought sheets are typically produced by cold rolling. Wires for structural welding are drawn down from hot rolled coil feedstock.

Many intermediate solution anneals are necessary during the cold rolling and cold drawing operations, because cobalt alloys exhibit high work hardening rates. Solution annealing is normally carried out at temperatures in excess of 1100 °C and is followed by rapid cooling to prevent the precipitation of grain boundary carbides.

With regard to castings of the (high-carbon) cobalt wear alloys, these are typically made using the investment casting process or the resin shell casting process. Normally, small quantities of remelt bar or remelt ingot are taken from larger arc-melted heats of material for this purpose. Investment cast molds, made by the lost-wax process, are usually clamped to high-frequency furnaces containing small quantities of molten remelt material, and rolled over to produce the castings. Pouring temperatures are in the vicinity of 1650 °C.

Weld overlay consumables of the same materials are produced by several methods, depending on the type of consumable desired. Straight lengths of wire, for gas tungsten arc (TIG) and shielded metal arc welding (with a suitable flux coating), are typically made by continuous casting. Coiled wires for gas metal arc (MIG) welding are made by wrapping the alloying ingredients in sheaths of cobalt (actually a cobalt–iron alloy, made by powder metallurgical techniques). Powders for plasma transferred arc welding are made by gas atomizing.

Bibliography

Massalski T B, Okamoto H, Subramanian P R, Kacprzak L 1990 *Binary Alloy Phase Diagrams*, 2nd edn. ASM International, Materials Park, OH, Vol. 2, pp. 1214–1215

D. Klarstrom and P. Crook
Haynes International Inc., Kokomo, Indiana, USA

Composite Materials, Microstructural Design of

Any two materials could, in principle, be combined to make a composite (Fig. 1), and they might be mixed in many geometries (Fig. 2). This article summarizes a scheme for identifying the component materials that could be used to make composites with potentially attractive properties. Three concepts are used. The first is that of performance indices, which isolate the combination of material properties that maximize performance; the second is that of materials-selection charts, onto which both material properties and performance indices can be plotted; and the third is the use of bounds to define the envelope of properties accessible to a given composite system (Ashby 1993).

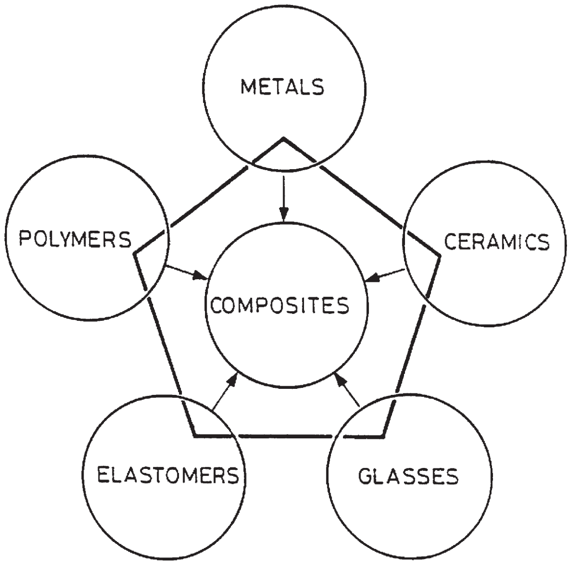


Figure 1
The material classes from which composites are made.

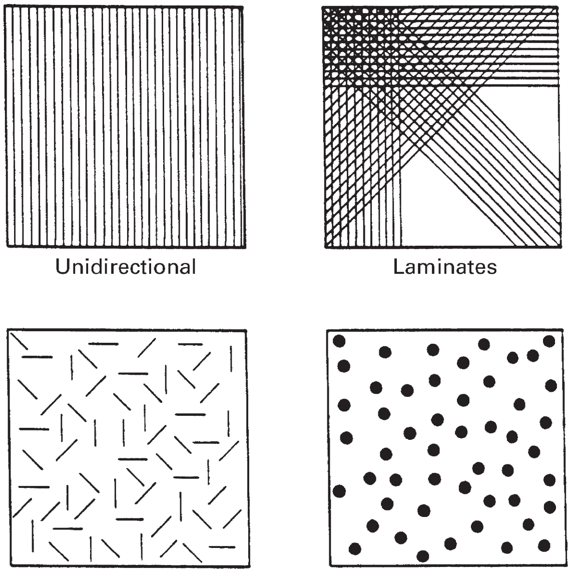

Unidirectional
Laminates
Chopped-fiber
Particulates

Figure 2
Unidirectional fibrous composites, laminated composites, chopped-fiber and particulate composites.

1. Indices, Selection Charts, and Property Bounds

A performance index is a property or group of properties that measures the effectiveness of a material in performing a given function (Ashby 1999). The best material for a light, stiff tie (a tensile member) is that with the greatest specific stiffness, E/ρ, but it is not always so straightforward. The lightest beam (a member loaded in bending) with a prescribed stiffness is that made of the material with the greatest value of $E^{1/2}/\rho$; the best springs are those made of materials with the highest values of σ_f^2/E, where σ_f is the yield or fracture stress; the precision device that is least distorted by heat is that made of the material with the highest value of λ/α, where λ is the thermal conductivity and α is the expansion coefficient. The performance indices are $E^{1/2}/\rho$, σ_f^2/E and λ/α. The best choice of material is that with the largest value of the appropriate index. The indices provide a set of pointers; they direct the composite-developer towards material combinations that, potentially, offer something new.

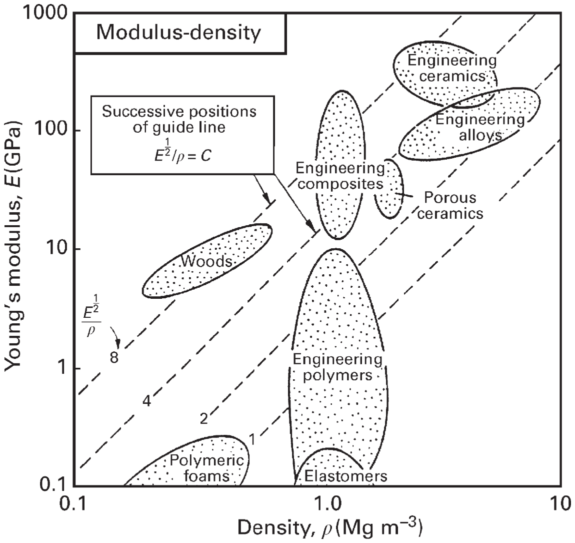


Figure 3
A schematic selection chart: modulus E is plotted against density ρ.

Table 1
Design limiting properties.

Property	Symbol and usual units
Density	ρ (Mg m^{-3})
Young's modulus	E (GPA)
Strength	σ_f (MPa, sometines GPa)
Toughness; or fracture toughness	J_{1c} (kJ m^{-2}); K_{1c} (MPam$^{1/2}$)
Linear thermal expansion coefficient	α (K^{-1})
Specific heat	C_p (J kg^{-1}K^{-1})
Thermal conductivity	λ (W m^{-1}K^{-1})
Thermal diffusivity	a (m^2 s^{-1})

The indices, frequently, combine two or more properties. This suggests the idea of constructing charts on which one property is plotted against another in such a way that both the properties and their combinations can be examined. Figure 3 illustrates this. The axes are Young's modulus E and density ρ. The scales are logarithmic, and span a range so wide that almost all engineering materials are included. The index E/ρ appears on these axes as a family of parallel lines of slope 1; the index $E^{1/2}/\rho$ appears as a family of slope 2. Other charts give optimal selection via other indices. Charts and indices have been fully described by Ashby (1999).

On a macroscopic scale a composite behaves like a homogenous solid with its own set of thermomechanical properties. Calculating these precisely is difficult. It is much easier to bracket them by bounds or limits: upper and lower values between which the properties lie. For most properties, this can be done accurately enough to identify interesting possibilities.

Seven mechanical and thermal properties are of primary interest in assessing the potential of a new composite: density, modulus, strength, toughness, thermal conductivity, expansion coefficient, and heat capacity (Table 1); others, like fracture toughness and thermal diffusivity, are calculated from them. In this section we assemble bounds or limits for material properties. The term "bound" will be used to describe a rigorous boundary, one that the value of the property *cannot*—subject to certain assumptions—exceed or fall below. However, it is not always possible to derive bounds; then the best that can be done is to derive "limits" outside which it is *unlikely* that the value of the property will lie. In what follows, density and specific heat are calculated exactly; moduli, thermal expansion, thermal conductivity and diffusivity are bracketed by bounds, but strength and toughness can only be enclosed by limits. As far as possible, all are based on micromechanical modeling; their origins have been detailed by Hull (1981), Kelly and Macmillan (1986), Chamis (1987), and Ashby (1993). The important point is that the bounds or limits bracket the properties of all arrangements of matrix and reinforcement shown in Fig. 2; by using them we escape from the need to model individual geometries.

1.1 Density

When a volume fraction f of a reinforcement r (density ρ_r) is mixed with a volume fraction $(1-f)$ of a matrix m (density ρ_m) to form a composite with no residual porosity or void-space, the composite density is given exactly by a rule of mixtures (an arithmetic mean, weighted by volume fraction)

$$\rho_r = f\rho_r + (1-f)\rho_m \tag{1}$$

1.2 Modulus

The modulus of a composite is bracketed by the well-known Voigt and Reuss bounds:

$$E_r = fE_r + (1-f)E_m \tag{2}$$

and

$$E_r = \frac{E_m E_r}{fE_m + (1+f)E_r} \tag{3}$$

Here E_r is the Young's modulus of the reinforcement and E_m that of the matrix. Tighter bounds exist, but Eqns. (2) and (3) are sufficient for our purpose.

1.3 Strength

The upper bound is, as with modulus, a rule of mixtures

$$(\sigma_f)_u = f(\sigma_f)_r + (1-f)(\sigma_f)_m \tag{4}$$

where $(\sigma_f)_m$ is the strength of the matrix and $(\sigma_f)_r$ is that of the reinforcement. A lower limit is the yield strength of the matrix enhanced slightly by the plastic constraint imposed by the reinforcement

$$(\sigma_f)_l = (\sigma_f)_m 1 + \frac{1}{16}\left(\frac{f^{1/2}}{1-f^{1/2}}\right) \tag{5}$$

The bounds are wide, but they still allow important conclusions to be reached.

1.4 Specific Heat

The specific heats of solids at constant pressure, C_p, are almost the same as those at constant volume, C_v. If they were identical, the heat capacity per unit volume of a composite would, like the density, be given exactly by a rule-of-mixtures

$$\rho C_p = f\rho_r(C_p)_r + (1-f)\rho_m(C_p)_m \tag{6}$$

where $(C_p)_r$ is the specific heat of the reinforcement and $(C_p)_m$ is that of the matrix.

1.5 Thermal Expansion Coefficient

We use the approximate lower bounds of Levin

$$\alpha_L = \frac{E_r \alpha f + E_m \alpha_m (1-f)}{E_r f + E_m(1-f)} \tag{7}$$

and the upper bound of Schapery

$$\begin{aligned}\alpha_u = & f\alpha_r(1+\nu_r) + (1-f)\alpha_m(1+\nu_m) \\ & - \alpha_L[f\nu_r + (1-f)\nu_m]\end{aligned} \tag{8}$$

where α_r and α_m are the two expansion coefficients and ν_r and ν_m the Poisson's ratios.

1.6 Thermal Conductivity

A composite containing parallel continuous fibers has a conductivity, parallel to the fibers, given by a rule-of-mixtures

$$\lambda_u = f\lambda_r + (1-f)\lambda_m \tag{9}$$

This is an upper bound; in any other direction the conductivity is lower. The transverse conductivity of a parallel-fiber composite (again assuming good bonding and thermal contact) lies near the lower bound first derived by Maxwell

$$\lambda_L = \left(\frac{\lambda_r + 2\lambda_m - 2f(\lambda_m - \lambda_r)}{\lambda_r + 2\lambda_m + f(\lambda_m - \lambda_r)}\right) \tag{10}$$

Particulate composites, too, have a conductivity near this bound. Details are given by Ashby (1993), in which bounds and limits for thermal diffusivity, toughness, and fracture toughness are also discussed.

2. Applications: The Potential of Composite Systems

The last step is that of combining the indices, the charts and the bounds to design composites for specific applications. Two examples are developed, but it will be clear that the method is a general one, and can be extended further.

2.1 Composite Design for Stiffness at Minimum Weight

Consider, first, design of a composite for a light, stiff, beam of fixed section-shape, to be loaded in bending. The efficiency is measured by the index $M = E^{1/2}/\rho$. Imagine, as an example, that the beam is at present made of an aluminum alloy and that the alloy could be stiffened by incorporating particles or fibers of beryllium (Be) or of alumina (Al_2O_3) in it. Both are much stiffer than aluminum—that is, they have higher moduli.

Figure 4 is a small part of the E–ρ chart introduced in the last section. Three groups of materials are shown: aluminum and its alloys; beryllium; and a range of aluminas. Composites made from them have densities given exactly by Eqn. (1), and moduli that are bracketed by the bounds of Eqns. (2) and (3). Both of these moduli depend on volume fraction of reinforcement, and through this, on density. Upper and lower bounds for the modulus–density relationship can thus be plotted on the E–ρ chart using volume fraction f as a parameter, as shown in Fig. 4. Any composite made by combining aluminum with beryllium will have a modulus that lies somewhere in the Al–Be envelope; any made of aluminum and alumina will have a modulus contained in the envelope for Al–Al_2O_3. Fibrous reinforcement gives a longitudinal modulus (that parallel to the fibers) near the upper bound; particulate reinforcement or transversely loaded fibers give moduli near the lower one.

Superimposed on Fig. 4 is a grid showing the performance indices M—the quantity we wish to maximize. The bound-envelope for Al–Be composites extends almost normal to the grid, while that for Al–Al_2O_3 is, initially, parallel to the M grid: 30% of particulate Al_2O_3 gives almost no gain in M. The underlying reason is clear: both beryllium and Al_2O_3 increase the modulus, but only beryllium decreases the density; both indices are more sensitive to density than to modulus.

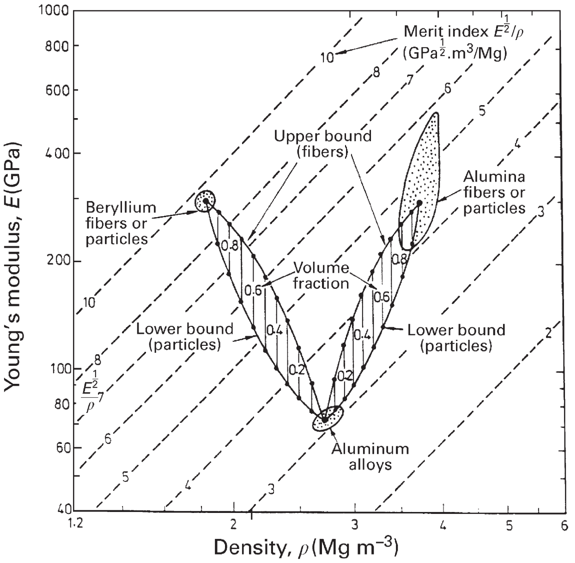


Figure 4
Part of a modulus–density space, showing aluminum alloys, beryllium and alumina. The moduli of the Al–Be and the Al–Al_2O_3 composites are bracketed by the bounds of Eqns. (1), (2), and (3). The Al–Be composite offers much greater gains in M than the Al_2O_3 composite.

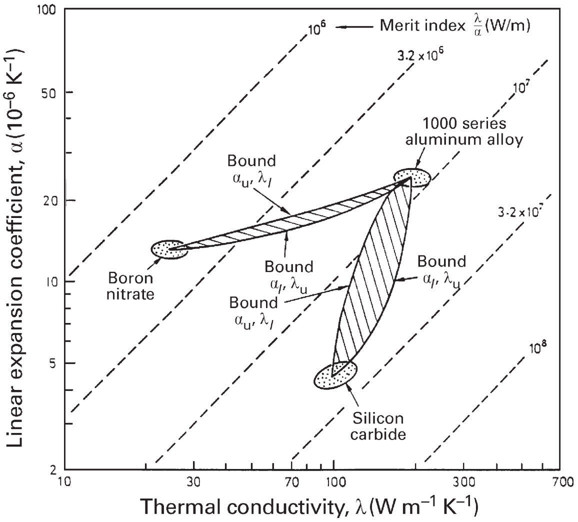


Figure 5
A part of expansion coefficient–thermal conductivity space, showing aluminum alloys, boron nitride and silicon carbide. The properties of Al–BN and Al–SiC composites are bracketed by the bounds described by Eqns. (7)–(10). The Al–SiC composites have better values of the index M (which increases towards the bottom right) than the Al–BN composites.

2.2 Composite Design for Specific Thermal Properties

Thermo-mechanical design involves the specific heat, C_p, the thermal expansion, α, the conductivity, λ, and the thermal diffusivity, $\alpha = \lambda(\rho C_p)^{-1}$. These composite properties are bounded by Eqns. (6)–(10). They are involved in a number of performance indices. One is the criterion for minimizing thermal distortion $M-\lambda/\alpha$.

Figure 5 shows a small part of the $\lambda-\alpha$ materials selection chart, with a grid of lines of the index M superimposed on it. Three groups of materials are shown, aluminum alloys, boron nitride (BN), and silicon carbide (SiC). The thermal properties of Al–BN and Al–SiC are bracketed by lines that show the bounding equations. The plot reveals immediately that SiC reinforcement in aluminum increases performance (as measured by M); reinforcement with BN decreases it. By systematically testing alternative combinations of material and reinforcement in this way, the pair that offers the greatest potential gain in performance can be identified.

3. Summary and Conclusions

Developing a new composite is a long and expensive business. It is helpful, before starting, to have an idea of what its strengths and weaknesses might be and where its applications might lie. This article outlines a procedure for doing this. It combines the ideas of performance indices, materials selection charts and bounds or limits for composite properties to identify the composites which, potentially, have attractive combinations of mechanical and thermal properties. Two examples of the method are given in the text. The same approach can be applied to a wide range of thermal and mechanical applications.

Bibliography

Ashby M F 1993 Criteria for selecting the components of composites. *Acta Metall. Mater.* **41**, 1313–35

Ashby M F 1999 *Materials Selection in Mechanical Design*, 2nd edn. Butterworth Heinemann, Oxford, UK

Chamis C C 1987 In: Weeton J W, Peters D M, Thomas K L (eds.) *Engineers Guide to Composite Materials*. American Society for Metals, Metals Park, OH, pp. 3-8–3-24

Hull D 1981 *An Introduction to Composite Materials*. Cambridge University Press, Cambridge, UK

Kelly A, Macmillan 1986 *Strong Solids*, 3rd edn. Oxford University Press, Oxford, UK

M. F. Ashby
University of Cambridge, UK

Copper Alloys: Alloy and Temper Designation

Copper and its alloys are classified by composition according to the Unified Numbering System (UNS) for metals and alloys. This system has been adapted directly from the one constructed many years ago by the Copper Development Association (CDA) and has been incorporated by other engineering specifications systems (ASTM, ASME, and SAE). Copper alloys are designated with a five-digit number preceded by the letter "C" and separated into two broad groups: C10100 to C79900 for wrought alloys and C80100 to C99900 for casting alloys. The alloys are further broken down into alloy families based on the principal alloying elements involved. While there is little numerical correspondence, similar alloy families are found in both the wrought and casting alloys (Table 1). Examples of prominent wrought copper alloys are listed with properties in Table 2, showing the strength/electrical conductivity trade-offs available across the alloy spectrum.

The coppers (C10100–C15999) have a minimum of 99.3 wt.% copper and may have very small intentional additions, including oxygen, silver, phosphorus, zirconium, or tin, for deoxidization or for softening resistance, while preserving the high conductivity of pure copper. The more highly alloyed coppers are grouped according to their principal alloy additive, such as zinc (brasses), tin (phosphor bronzes), aluminum (aluminum bronzes), nickel (cupronickels), etc.

The high-copper alloy group (C16200–C19999), with copper content >94 wt.%, is a catchall for a variety of alloying elements used in relatively small amounts for increasing the hardness and strength with dispersion and precipitation hardening second phases, while still maintaining good levels of conductivity and formability. Selected examples of the variety of hardening phases found amongst the alloys in this high-copper alloy group are the iron particles in alloy C19400, chromium precipitates in alloy C18200, and the pre-precipitates in the premier age-hardening beryllium copper alloys, C17200 and C17500, for example.

The copper zinc alloys (brasses) form a family of solid solution alloys (C20500–C26000) offering good formability at good strength and, where color is important, pleasing appearance. The properties of the base copper–zinc brasses are improved with other alloy additions: up to 3 wt.% lead to form dispersed particles to promote machinability, tin for solution strengthening, and tin, arsenic, or antimony for corrosion resistance.

Phosphor bronze alloys (C50500–C52400) are solid solution alloys that offer excellent combinations of strength, ductility, and corrosion resistance, to which lead can also be added (C54400) for machinability and bearing properties. Aluminum bronze alloys (C60800–C64210), containing 2–15 wt.% aluminum, and silicon bronze alloys (C64700–C66100), containing 0.5–4.0 wt.% silicon, offer good strength via the high work hardening rates effected by these two solid solution strengthening elements along with good corrosion resistance. An addition of 1–5 wt.% iron is used in aluminum bronze alloys (C63000, for example), for the dispersion hardening and grain size control provided by the resulting iron-rich particles.

Because of their good corrosion resistance the cupronickel alloys (C70100–C72950) are used in heat exchangers and US coinage. The copper–nickel–zinc family of alloys (C73500–C79800), traditionally called "nickel silvers," utilizes solid solution strengthening and work hardening to offer a good combination of strength and formability, with good corrosion and tarnish resistance.

The various copper alloys are manufactured to finish in a range of tempers from dead soft or fully annealed for maximum formability to heavily worked or age-hardened conditions for maximum strength and hardness, whether the product form is rod, wire, plate, strip, or tube, or a casting. The system for designating the material or product condition is defined in ASTM B 601, *Standard Practice for Temper Designations for*

Table 1
Copper Development Association and ASTM classification of copper alloys.

Wrought alloys		Cast alloys	
Cu and high-Cu alloys	C10000	Cu and high-Cu alloys	C80000 to 83000
Zn brasses	C20000	Zn, Zn–Sn, Zn–Pb brasses	C83300 to 85800
Zn–Pb brasses	C30000	Mn bronzes	C86100 to 86800
Zn–Sn brasses	C40000	Si bronzes and brasses	C87200 to 87900
Sn bronzes	C50000	Sn and Sn–Pb bronzes	C90200 to 94500
Al, Mn, Si bronzes	C60000	Ni–Sn bronzes	C94700 to 94900
Cu–Ni, Cu–Ni–Zn alloys	C70000	Al bronzes	C95200 to 95800
		Ni–Cu alloys	C96200 to 96600
		Zn–Ni–Cu alloys	C97300 to 97800
		Pb–Cu alloys	C98200 to 98800

Table 2
Compositions and properties of selected copper alloys representing wrought copper alloy families.

Alloy, Composition	Electrical conductivity (MS m^{-1}, %IACS)	Temper	Tensile strength (MPa)	Yield Strength (MPa)
Coppers (C10100–C15999)				
C10200 99.95 Cu	58, 101	H08	372	358
C12200 Cu–0.02P	49, 85	H08	372	358
High-copper alloys (C16200–C19999)				
C17200 Cu–1.9Be–0.4(Ni + Co)	13, 22	TM08	1240	1120
C18200 Cu–0.9Cr	46, 80	TH02	462	406
C19400 Cu–2.4Fe–0.13Zn–0.04P	38, 65	H08	503	482
Cu–Zn brasses (C21000–C28000)				
C23000 Cu–15Zn	21, 37	H04	469	420
C26000 Cu–30Zn	16, 28	H04	524	496
Cu–Zn–Pb brasses (C31200–C38500)				
C35300 Cu–36Zn–2Pb	15, 26	H04	503	462
Cu–Zn–Sn brasses (C40400–C48600)				
C42500 Cu–9.5Zn–2Sn–0.2P	16, 28	H04	524	496
C44300 Cu–28Zn–1Sn–0.04As	14, 25	WM02	503	434
Phosphor bronzes (C50100–C54200)				
C51000 Cu–5Sn–0.2P	8.7, 15	H08	706	689
C52400 Cu–10Sn–0.2P	6.4, 11	H08	772	734
C54400 Cu–4Sn–4Zn–4Pb–0.2P	11, 19	H04	548	531
Aluminum bronzes (C60800–C64210)				
C63000 Cu–10Al–5Ni–3Fe	4.1, 7	H02	813	510
C63800 Cu–2.8Al–1.8Si–0.4Co	5.8, 10	H04	827	723
Silicon bronzes (C64700–66100)				
C65500 Cu–3Si–1Mn C	4.1, 7	H04	648	393
Modified Cu–Zn brasses				
C67500 Cu–39Zn–1.4Fe–1Sn–0.2Mn	14, 24	H02	579	413
C68800 Cu–23Zn–3Al-0.4Co	10, 18	H04	758	690
C69400 Cu–14.5Zn–4Si	3.6, 6.2	O60	586	296
Cupronickels (C70100–C72950)				
C70250 Cu–3.0Ni–0.65Si-0.15Mg	23, 40	TM02	703	627
C70600 Cu–10Ni-1.4Fe	5.2, 9	H55,H01	379	310
C71500 Cu–30Ni–0.5Fe	2.7, 4.6	H55,H01	482	413
Nickel silvers (C73500–C79800)				
C74500 Cu–25Zn–10Ni	5.2, 9	H04	593	510
C77000 Cu–27Zn–18Ni	3.2, 5.5	H08	785	770

Copper and Copper Alloys Wrought and Cast. A few examples are given in Table 3, where it is seen that the designations are product form related, indicating processing only in a generic sense. The alloy supplier will apply the specific processing parameters needed to meet the property specifications, usually tensile strength, listed in industrial specifications for a given temper. For example, in making a strip product in the HR04 (Hard Relief Annealed) temper, the alloy producer will start with the fully annealed condition, cold roll the correct reduction in thickness to finish gauge, and apply the proper relief annealing treatment required to meet the strength and ductility parameters specified for the HR04 temper product.

There are many more temper designations applied to wrought products versus cast products because of the greater range of properties that are readily available through combination of processing steps, including working before or after age-hardening treatments. Nonetheless, cast products may require heat treatments for homogenization or solutionization, as well as for age-hardening or stress relieving. Copper alloy suppliers provide many alloys in the so-called mill hardened temper, in which the material has already been given the solutionizing and aging thermal treatments during processing, thus avoiding the need for an age-hardening heat treatment of the final fabricated part.

Table 3
Examples of temper designations used for copper and copper alloys, from ASTM B 601.

Temper designation	Temper names or material condition
Annealed tempers	
O10	Cast and annealed
O20	Hot forged and annealed
O60	Soft annealed
O025	Average grain size: 0.025 mm
Cold worked tempers	
H01	$\frac{1}{4}$ Hard
H02	$\frac{1}{2}$ Hard
H04	Hard
H08	Spring
H14	Super spring
H55	Light drawn; light cold rolled
H86	Hard drawn electrical wire
Cold worked and stress relieved tempers	
HR02	H02 and stress relieved
HR08	H08 and stress relieved
Precipitation hardened tempers	
TB00	Solution heat treated
TF00	TB00 and age hardened
TH02	TB00 cold worked and age hardened
TL02	TF00 cold worked to $\frac{1}{2}$ Hard
TM00, TM02, TM08, etc.	Mill hardened tempers
TR02	TL02 and stress relieved
As manufactured tempers	
M01	As sand cast
M04	As pressure die cast
WM02	Tubing welded from H02 strip

Bibliography

ASTM B 601 1999 *Standard Practice for Temper Designations for Copper and Copper Alloys Wrought and Cast*. American Society for Testing and Materials *Annual Book of ASTM Standards*. ASTM, Philadelphia, PA, Vol. 02.01, pp. 715–19.

Copper Development Association 1985 *Standards Handbook*, 8th edn, Part 2—Alloy data wrought copper and copper alloy mill products, CDA, New York

Copper Development Association 1993 *Copper Casting Alloys: Alloy Selection Guide, Properties, Specifications, Casting Design*. CDA, New York

Copper Development Association 1998 *Standard Designations for Wrought and Cast Copper and Copper Alloys, Application Data Sheet*. CDA, New York

Metals Handbook 1990 Properties and Selection: Nonferrous Alloys and Special-purpose Materials. ASM International, Materials Park, OH, 10th edn., Vol. 2, pp. 216–427

Peters D, Kundig K 1994 Selecting coppers and copper alloys–Part I: Wrought products. *Adv. Mater. Process.* **151** (2), 26–32

Peters D, Kundig K 1994 Selecting coppers and copper alloys–Part II: Cast products. *Adv. Mater. Process.* **151** (6), 20–6

R. N. Caron
Olin Corporation, New Haven, Connecticut, USA

Crystal Engineering

Crystal engineering is a general term covering the processes of purposeful design of functional three-dimensional crystal structures from molecular-scale components, and of crystals with particular morphology and size. These processes are of fundamental importance to the development of new materials, since a vast array of properties (e.g., optical and nonlinear optical properties, electrical conductivity, magnetism, solubility, photochemical stability, and porosity) are determined in part by crystal structure. Practical applications of crystalline materials are also strongly influenced by the ability to grow crystals of suitable size and morphology. In some cases large single crystals are preferred while in other situations small regularly-shaped crystals of a product are required, e.g., to produce free-flowing crystalline products in industry. Two scientific developments which have contributed to rapid progress in both of these aspects are (i) the availability of over 140,000 crystal structures of organic materials in the computer-searchable Cambridge Structural Database (CSD) (Allen *et al.* 1983, 1991, 1994, and (ii) increased computational power which permits modeling of crystal structures and calculation of surface energies for growing crystal faces as well as analysis of the significance of intermolecular interactions.

1. Supramolecular Chemistry

The first principle of crystal engineering is the analysis of groups of related crystal structures to identify important substructural units (also termed motifs, patterns, building blocks, and synthons) which occur repeatedly in many known crystal structures and can thus be relied upon to be influential in steering molecules into desired packing arrangements in new materials. Lehn (1995) first pointed out that "just as there is a field of *molecular chemistry* based on the covalent bond, there is a field of *supramolecular chemistry*, the chemistry of molecular assemblies and the intermolecular bond." Desiraju (1989, 1995) has drawn an elegant parallel between the principles of organic synthesis in molecular chemistry, as described by Corey (1988), and those of crystal engineering in supramolecular chemistry. In both molecular and supramolecular chemistry retrosynthetic methods, using a "disconnection" approach to dissect a complex molecule or crystal structure into smaller subunits connected by available synthons, are in principle possible.

2. Supramolecular Synthons

Supramolecular synthons have been identified from systematic searches of the CSD made by defining distances and geometrical attributes of nonbonded interactions between defined atom types or defined groups of atoms with particular connectivity. Scatter plots of distances and directions between nearest nonbonded atoms reveal persistent interaction patterns (supramolecular synthons), many of which are listed in Desiraju (1995). They include systems analogous to single, double, and triple bonds and both weak and strong interactions (e.g., Fig. 1). Their energetic significance may be seen by comparing the typical interaction energies for van der Waals forces $<8\,\mathrm{kJmol^{-1}}$) and hydrogen bonds (weak hydrogen bonds such as $\mathrm{CH\cdots O}$ and $\mathrm{OH}\cdots\pi$, 2–$20\,\mathrm{kJ\,mol^{-1}}$; $\mathrm{OH\cdots O}$, 20–$40\,\mathrm{kJ\,mol^{-1}}$; $\mathrm{XH\cdots X}$ where $\mathrm{X}=$ halogen, phosphorus, nitrogen, etc., 40–$190\,\mathrm{kJ\,mol^{-1}}$ depending on degree of ionic character) with a typical C—C covalent bond energy of about $300\,\mathrm{kJmol^{-1}}$. Thus the energy gained, particularly where two molecules link using two or more strong hydrogen bonds, in structures optimizing such interactions is indeed comparable to that of weak covalent bonding. The term "synthon" is therefore realistic. These synthons may be used to design specific supramolecular architectures. For example, the hydrogen-bonded carboxylic acid dimer synthon may be used to form hexagonal sheet structures from benzene 1,3,5-tricarboxylic acid (trimesic acid) (although these sheets are not planar and form a complex three-dimensional interpenetrating network of four separate sheets—see Childs) and three-dimensional diamond-like structures from adamantane 1,3,5,7-tetracarboxylic acid (Fig. 2). Inorganic and organometallic compounds can also be incorporated into such networks (Braga and Grepioni 1996, Braga. For example trimesic acid (TMA) forms a hydrated anionic framework around stacks of cobalticinium cations in the compound $[(\mathrm{C_5H_5})_2\mathrm{Co}][(\mathrm{H_3TMA})(\mathrm{H_2TMA})]\cdot 2\mathrm{H_2O}$. Weaker hydrogen bonds (e.g., $\mathrm{CH\cdots C}$) can also exert significant influences on crystal structures; the large number of these contacts in common structures amplifies their overall contribution to the total lattice energy to a significant level despite the weakness of the individual interactions. Thus, the occurrence of "herring-bone" stacking in crystals of aromatic hydrocarbons has been correlated with the H:C ratio in the molecule, molecules with high ratios favoring this type of packing which places peripheral hydrogen atoms of one molecule close to carbon atoms of inclined molecules in neighboring stacks (Desiraju and Gavezzotti 1989).

Although strong hydrogen-bonded interactions form the largest group of supramolecular synthons, halogen atom interactions with nitro, cyano, amino, and imino groups and other halogen atoms are also common, and $\mathrm{S\cdots S}$ interactions play a significant role in determining the stacking arrangements in organic metals involving donors such as TTF (tetrathiafulvalene and BEDT–TTF (tris(ethylenedithio) tetrathiafulvalene. These interactions have as their origin not only dispersion forces but also the nonspherical shape of the atoms arising from the presence of directed lone pairs (Price and Stone 1984). The presence of specific attractive forces between halogens and other atoms has been demonstrated by showing that the number of nearest-neighbor contacts to halogen atoms in crystals is larger than statistically expected from considerations of atomic size alone (Desiraju and Parthasarathy 1989) and gas-phase molecular beam experiments have demonstrated that such interactions are not confined to the solid state.

The interactions of neighboring π-electron systems (π–π interactions) have also been shown to be optimized by specific overlap geometries. This has been explained in terms of a simple electrostatic model by Hunter and Sanders (1990), in which the p-electrons of the π-system are considered as fractional negative charges above and below the plane of positively charged atom cores. Clearly, from this model, stacks of directly superposed molecules are unstable due to repulsion of the adjacent negatively charged regions, whereas T-shaped alignment or stacks with rings displaced by half a ring diameter are favorable due to attraction of the negative and positive regions of neighboring molecules.

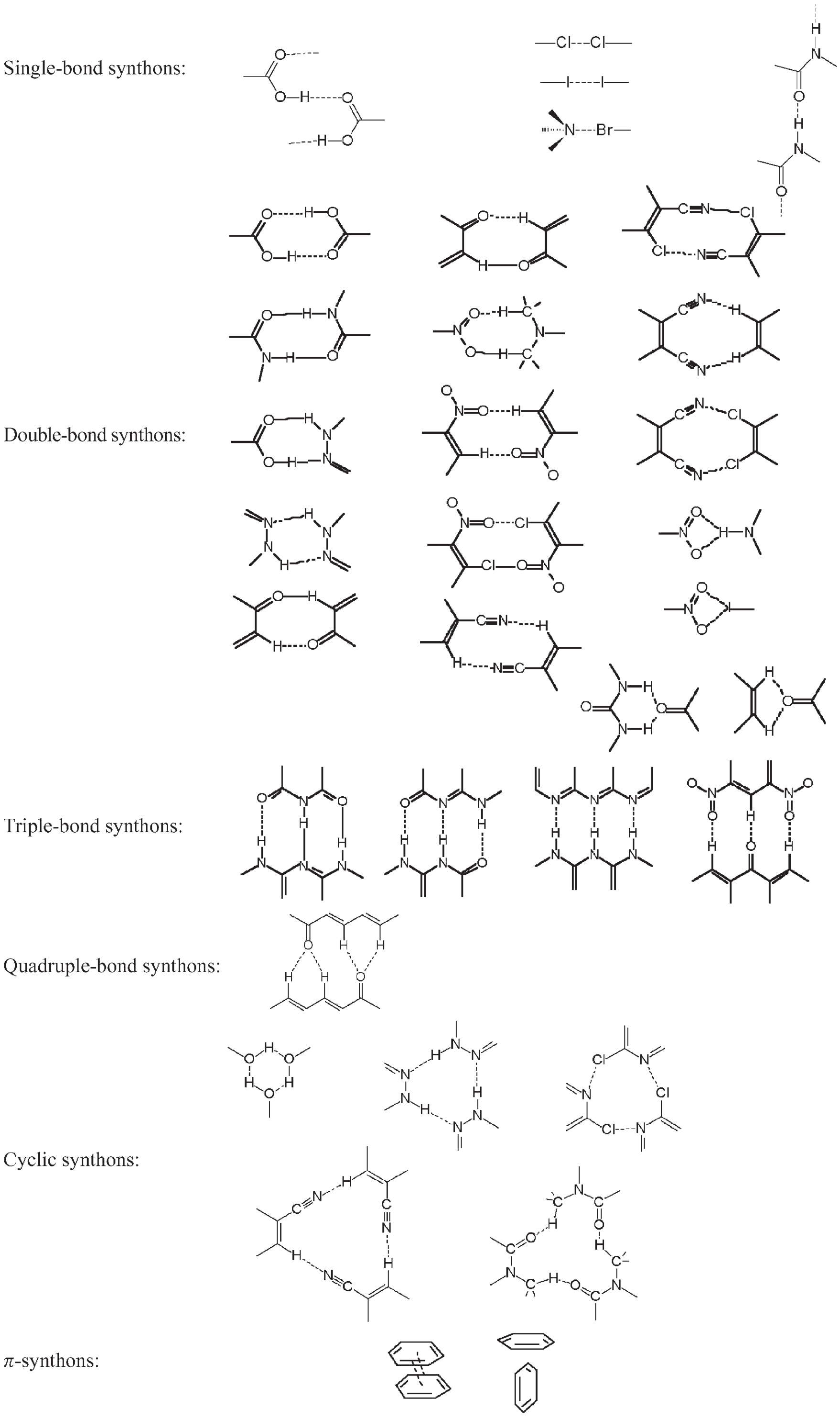

Single-bond synthons:
Double-bond synthons:
Triple-bond synthons:
Quadruple-bond synthons:
Cyclic synthons:
π-synthons:

Figure 1
Supramolecular synthons.

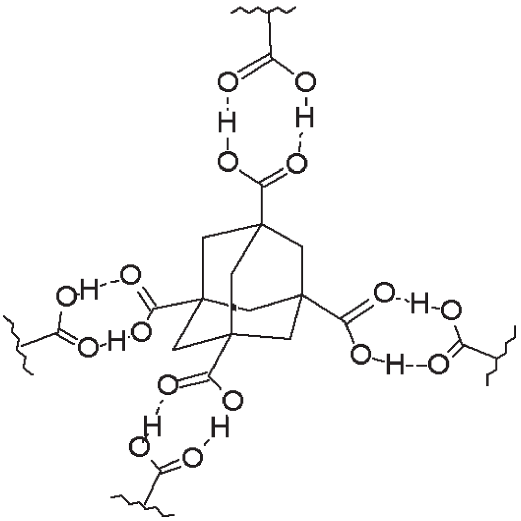


Figure 2
Adamantane 1,3,5,7-tetracarboxylic acid.

In crystals composed of π-electron donors and acceptors, charge-transfer interactions also influence the observed molecular orientations (Wright 1995). Two classes of π–π^* charge-transfer systems may be distinguished, involving electronic ground states with small and large degrees of charge transfer, respectively. In the former, donor and acceptor molecules alternate in stacked arrangements, with neighboring stacks often arranged to form planar sheets of molecules. The relative orientation of donor and acceptor within a stack is frequently close to that which maximizes the interaction energy of all the filled π orbitals of the donor with all the vacant π^* orbitals of the acceptor, although this is not always the orientation predicted from simplified frontier orbital models involving only interaction of donor HOMO and acceptor LUMO. In systems where the degree of charge transfer in the ground state is large, homosoric structures with stacks composed entirely of donors or acceptors are often found. These have been rationalized in terms of solution equilibria during crystal growth, with low equilibrium concentrations of donor$^+$ (which is an electron acceptor, and thus complexes with neutral donor) and acceptor$^-$ (which is an electron donor, and thus complexes with neutral acceptor). These D^+–D and A^-–A complexes serve as templates for subsequent stack growth, and illustrate the significance of solution equilibria in crystal engineering.

3. *Control of Crystal Structure and Morphology*

Supramolecular synthons can in principle be used to design crystals in much the same way as conventional molecular synthons in organic chemistry are used in retrosynthetic analysis with a disconnection approach. The desired final structure is disconnected into smaller structural subunits, which can be re-assembled using known supramolecular synthons, and the process is continued until a convenient starting molecule is identified. However, in practice this is very difficult due to the numerous complexities of organic crystals (Gavezzotti 1994): it is currently impossible to design a molecule to form a crystalline solid with reliably predictable structure for a given materials application. The retrosynthetic approach using supramolecular synthons should therefore be regarded as a useful starting point rather than an infallible procedure. A major complicating factor is the occurrence of polymorphism. (Although less than 5% of the compounds in the CSD are known to be polymorphic, this is likely to be a significant underestimate since many structures are determined using only one crystal specimen.) Frequently different polymorphs have very similar lattice energies, especially since polymorphs with smaller sublimation enthalpies are often less dense and thus have higher vibrational entropy. In such cases, several polymorphic forms can crystallize simultaneously from the same solution.

According to Ostwald's "Rule of stages" crystallization initially proceeds so as to incur the smallest possible change in free energy, i.e., the most soluble metastable forms will form initially in preference to the final most stable polymorph (although once seeds of the latter are available its formation is favored—see Dunitz and Bernstein (1995)). This strong influence of seed crystallites can be a serious problem where crystals of a single polymorph must be obtained (e.g., in the pharmaceutical industry), particularly since the critical size of a seed nucleus is estimated to range from a few tens to a few millions of molecules. Even for the upper end of this range, 10^{-6}g of the material (containing about 10^{16} molecules of relative molecular mass 100) could make at least 10^9 nuclei, so accidental seeding becomes virtually inevitable. Hence a development in crystal engineering has been the understanding of the importance of the surface of the growth nucleus in relation to specific interactions with solvent molecules (Davey 1982) or deliberately added tailored growth inhibitors (Weissbuch *et al.* 1991; Davey *et al.* 1997).

Growth inhibitors bind to a specific crystal face using one or more supramolecular synthons, but their reverse side no longer provides the necessary functionality for further binding. Since crystals elongate along the direction of fastest growth, inhibitors enlarge the faces to which they are bound (i.e., growth in the plane of the face is faster than growth perpendicular to it), changing crystal morphology in predictable ways. A simple example of this is the modification of benzamide crystals, which normally grow as {001} plates elongated along *b*, with doubly hydrogen-bonded amide dimers linked together by further double NH···O interactions along the *b* axis.

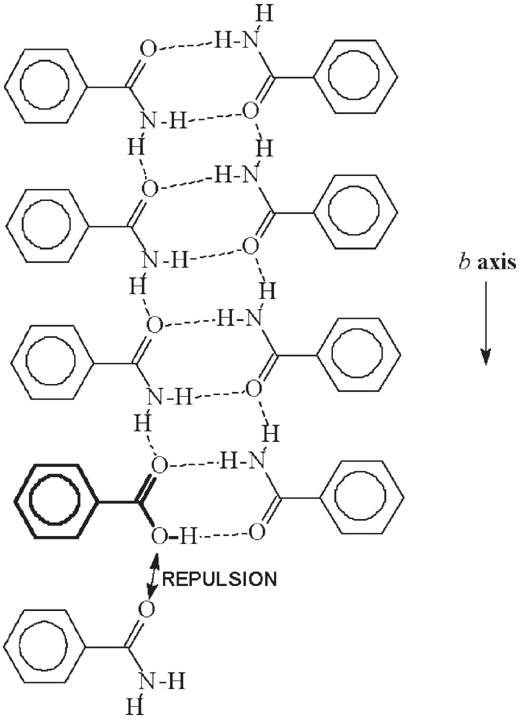


Figure 3
Benzoic acid as a growth inhibitor for benzamide.

Addition of small amounts of benzoic acid replaces the NH_2 groups by OH, leaving no spare hydrogen for sustaining the hydrogen bonding along *b* (Fig. 3).

The replacement of NH_2 by OH has been estimated to reduce the local interaction energy at the growth site by $40\,\mathrm{kJ\,mol^{-1}}$ so that the crystals then elongate along *a*. Chiral growth inhibitors have also been used successfully to inhibit crystal growth of one chirality, producing crystals of the noninhibited form. Since growth inhibition is a surface process, only small amounts of chirally pure inhibitor are required to produce large amounts of the resolved crystals. A related principle has also been used to grow chiral crystals of amino acids and other materials at interfaces with Langmuir–Blodgett films of chirally pure amino acids such as (*S*)-α-aminooctanoic acid. Using very small quantities of chirally pure material as monolayer templates, substantial quantities of optically pure crystals, whose growth has been otherwise entirely unconstrained in solution, can be obtained.

4. *Control of Twinning*

Certain faces of crystals incorporating hydrogen-bonded dimers can be generated either by cleaving between neighboring dimer pairs or by separating the two molecules in each surface dimer pair. The $\{10\bar{2}\}$ face of saccharin is one such example, and yields a surface functionality which depends on the growth solvent. Modeling of interactions of monomer and dimer with solvents showed that ethanol could hydrogen bond with both monomer and dimer species, whereas acetone could only hydrogen bond with the monomer. Hence the monomer surface, generated by separating the two molecules in dimer pairs at the cleavage plane, is favored in acetone solution but not in ethanol. This has been shown to affect the probability of developing twinned crystals. Modeling shows that dimers can add in two alternative orientations on to the monomer surface, but in only one way on to the dimer surface. This explains the experimental observation that twinned crystals form in acetone but not in ethanol (Williams-Seton *et al.* 1999). Such detailed understanding of twin growth processes can have significant industrial benefits.

5. *Conclusions*

The above examples of crystal growth illustrate the complexities that can occur due to competing interactions within a crystal. Thus, benzamide crystal growth is influenced both by the interactions forming benzamide hydrogen-bonded dimers and by the hydrogen-bonding between adjacent dimers, while the retention or cleavage of such dimers at a crystal surface in saccharin strongly influences the probability of crystal twinning. Since many functional organic molecules have a variety of features potentially capable of participating in different supramolecular synthons (e.g., aromatic rings, hydrogen-bonding substituents such as amine or carboxylic acid groups, and other interacting substituents such as cyano, halogen, or nitro groups, etc.) there will in general be a competition between the structural arrangements required to maximize their diverse interactions. Even in simple cases such as centrosymmetric carboxylic acid dimers, alternatives such as the chain structure can compete. In many cases this competition is further complicated by binding to solvent molecules (20% of the structures in the CSD contain guest molecules, usually solvents). In view of these complications, it is not surprising that computer prediction *ab initio* of the observed crystal structure of a designed functional molecule is not yet possible (Gdanitz 1998). (This is paralleled on the experimental side by the not infrequent observation that new compounds cannot be isolated as good-sized single crystals.) Indeed, full structure prediction will probably require computation of both thermodynamic and kinetic pathways for the whole process from solution disorder to crystal order, including liquid structuring and nucleation, crystal growth, solvent interactions, and effects of molecular motion and disorder within crystals. Despite these complications, the growing understanding of crystal engineering principles is vital to the future rational design of molecular materials for device applications, and the combination of statistical analysis of structural databases with increasingly sophisticated computer modeling techniques may be expected to generate continued progress in this area for the foreseeable future.

Bibliography

Allen F H, Davies J E, Galloy J J, Johnson O, Kennard O, Macrae C F, Watson D G 1991 The development of version 3 and version 4 of the Cambridge Structural Database system. *J. Chem. Inf. Comput. Sci.* **31**, 187–204

Allen F H, Kennard O, Taylor R 1983 Systematic analysis of structural data as a research technique in organic chemistry. *Acc. Chem. Res.* **16**, 146–53

Allen F H, Kennard O, Watson D G 1994 In: Bürgi H-B, Dunitz J D (eds.) *Structure Correlation*. VCH, Weinheim, Vol. 1, p. 74

Braga D *http://calullo.ciam.unibo.it/index.html.braga*

Braga D, Grepioni F 1996 Intermolecular interactions and supramolecular organisation in organometallic solids. *Chem. Commun.*, 571–8

Cambridge Structural Database *http://www.ccdc.cam.ac.uk/prods/csd*

Childs S http://euch3i.chem.emory.edu/~childs/supramolecular

Colloids, Crystals and Interfaces research Group, UMIST. *http://www.ce.umist.ac.uk/Research/ccio.asp*

Corey E J 1988 Retrosynthetic thinking—essentials and examples. *Chem. Soc. Rev.* **17**, 111–33

Davey R J 1982 Solvent effects in crystallization process. In: Kaldis E (ed.) *Current Topics in Materials Science*. North-Holland, New York, Vol. 8. Chap. 6

Davey R J, Blagden N, Potts G D, Docherty R 1997 Polymorphism in molecular crystals: stabilization of a metastable form by conformational mimicry. *J. Am. Chem. Soc.* **119**, 1767–72

Desiraju G R 1989 *Crystal Engineering*. Elsevier, Amsterdam

Desiraju G R 1995 Supramolecular synthons in crystal engineering—a new organic synthesis. *Angew. Chem. Int. Ed. Engl.* **34**, 2311–27

Desiraju G R, Gavezzotti A 1989 Crystal structures of polynuclear aromatic hydrocarbons – classification, rationalisation and prediction from molecular structure. *Acta Crystallogr.* **B45**, 473–82

Desiraju G R, Parthasarathy R 1989 The nature of halogen–halogen interactions—are short halogen contacts due to specific attractive forces or due to close packing nonspherical atoms? *J. Am. Chem. Soc.* **111**, 8725–6

Dunitz J D, Bernstein J 1995 Disappearing polymorphs. *Acc. Chem. Res.* **28**, 193–200

Gavezzotti A 1994 Are crystal structures predictable? *Acc. Chem. Res.* **27**, 309–14

Gdanitz R J 1998 *Ab initio* prediction of molecular crystal structures. *Curr. Opin. Solid State Mater. Sci.* **3**, 414–8

Hunter C A, Sanders J K M 1990 The nature of π–π interactions. *J. Am. Chem. Soc.* **112**, 5525–34

Lehn J-M 1995 *Supramolecular Chemistry: Concepts and Perspectives*. VCH, Weinheim

Price S L, Stone A J 1984 A six-site intermolecular potential scheme for the azabenzene molecules, derived from crystal structure analysis. *Mol. Phys.* **51**, 569–83

Weissbuch I, Addadi L, Lahav M, Leiserowitz L 1991 Molecular recognition at crystal interfaces. *Science* **253**, 637–45

Williams-Seton L, Davey R J, Lieberman H F 1999 Solution chemistry and twinning in saccharin crystals: a combined probe for the structure and functionality of the crystal–fluid interface. *J. Am. Chem. Soc.* **121**, 4563–7

Wright J D 1995 *Molecular Crystals*, 2nd edn. Cambridge University Press, Cambridge, pp. 34–43

J. D. Wright
University of Kent, Canterbury, UK

Crystal Structures of the Elements

The crystal structures of most chemical elements resemble rather simple sphere packings at RTP (room temperature and pressure). There is a fairly clear trend as to what crystal structure is stable as a function of the position of an element in the periodic table (Table 1). Obviously, this is a result of the electronic structure of the solid elements and the resulting type of bonding. Nondirectional metallic or van der Waals bonding favors dense packing of almost spherical atoms. In the case of smooth, repulsive interatomic potentials, h.c.p. or c.c.p. (f.c.c.) structures are formed (Fig. 1). If the cohesion energy is governed by the electrostatic Madelung energy, b.c.c. structures are preferred. Solids with prevailing directional covalent bonding tend to form open complex structures.

Most of the elements undergo structural phase transitions as a function of temperature and/or pressure. The b.c.c. structure type, for instance, is very common for high-temperature (HT) phases owing to its higher vibrational entropy compared to h.c.p. or f.c.c. structures. Under compression, the electronic band structure and thereby the bonding properties change. This often leads first to the formation of low-symmetry complex structures which at higher pressure then transform into high-symmetry close-packed structures. In addition, the delocalization of bonding electrons under pressure reduces the differences between the chemical properties of the elements and their crystal structures. At ultra-high pressures (of the order of TPa) most of the nonmetallic elements are expected to undergo insulator–(semiconductor)–metal transitions. Polyatomic molecular structures dissociate under sufficiently high pressure. The rapid improvement of high-pressure (HP) techniques since the early 1980s has opened up extremely high pressure and temperature ranges (comparable to the conditions in the earth's core: ~7000 K and ~360 GPa) to structural studies. As a result, numerous new allotropes of the elements have been discovered (see Table 3).

The theoretical modeling of the stability of crystal structures is usually performed by self-consistent density-functional calculations using different types

Table 1
Periodic system of the chemical elements. Elements with dense-packed structure types *cI*2-W (b.c.c.), *cF*4-Cu (f.c.c.), *hP*2-Mg (h.c.p.), *hP*4-La (d.h.c.p.) at ambient conditions are marked. The rare gases show close-packed structures at low temperatures and/or high pressures. The other elements exhibit open structures in terms of the hard-sphere model.

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1 H			□ bcc	⬡ hcp	⊡ fcc	⏣ dhcp											2 He
3 □ Li	4 ⬡ Be											5 B	6 C	7 N	8 O	9 F	10 Ne
11 □ Na	12 ⬡ Mg											13 ⊡ Al	14 Si	15 P	16 S	17 Cl	18 Ar
19 □ K	20 ⊡ Ca	21 ⬡ Sc	22 ⬡ Ti	23 □ V	24 □ Cr	25 Mn	26 □ Fe	27 ⬡ Co	28 ⊡ Ni	29 ⊡ Cu	30 ⬡ Zn	31 Ga	32 Ge	33 As	34 Se	35 Br	36 Kr
37 □ Rb	38 ⊡ Sr	39 ⬡ Y	40 ⬡ Zr	41 □ Nb	42 □ Mo	43 ⬡ Tc	44 ⬡ Ru	45 ⊡ Rh	46 ⊡ Pd	47 ⊡ Ag	48 ⬡ Cd	49 In	50 Sn	51 Sb	52 Te	53 I	54 Xe
55 □ Cs	56 □ Ba	57 ⏣ La*	72 ⬡ Hf	73 □ Ta	74 □ W	75 ⬡ Re	76 ⬡ Os	77 ⊡ Ir	78 ⊡ Pt	79 ⊡ Au	80 Hg	81 ⬡ Tl	82 ⊡ Pb	83 Bi	84 Po	85 At	86 Rn
87 Fr	88 □ Ra	89 ⊡ Ac⁺															

*Lanthanide metals	58 ⏣ Ce	59 ⏣ Pr	60 ⏣ Nd	61 ⏣ Pm	62 Sm	63 □ Eu	64 ⬡ Gd	65 ⬡ Tb	66 ⬡ Dy	67 ⬡ Ho	68 ⬡ Er	69 ⬡ Tm	70 ⊡ Yb	71 ⬡ Lu
⁺Actinide metals	90 ⊡ Th	91 Pa	92 U	93 Np	94 Pu	95 ⏣ Am	96 ⏣ Cm	97 ⏣ Bk	98 ⏣ Cf	99 ⊡ Es	100 Fm	101 Md	102 No	103 Lr

of approximations. This means the full many-body problem is approximated by one-electron Hamiltonians with local exchange-correlation potentials. In some cases relativistic effects such as increase in electron mass with velocity and spin–orbit coupling also have to be considered. For nonzero temperatures the quantum molecular dynamics method of Car and Parinello (1989) is often used. Owing to the rapid development of all these techniques in combination with more powerful computers, the stability of most crystal structures of the elements and their phase transitions are now quite well understood. For a short introduction to *ab initio* calculations see, for instance, Carlsson and Meschter (1995) or Young (1991).

Most of the information given in this article is based on the works of Young (1991), Tonkov (1992, 1996), and Villars and Calvert (1991).

1. Nomenclature

For the shorthand characterization of crystal structures, the Pearson notation in combination with the prototype structure is used in accordance with IUPAC recommendations (Leigh 1990). A comparison of Pearson notation, prototype formula, space group, and *Strukturbericht* designation (A3 for *hP*2-Mg, for instance) for a large number of crystal structure types is given by Massalski (1990). The prototype structures are described in detail in Daams *et al.* (1991).

The Pearson notation consists of two letters and one number. The first letter denotes the crystal family and the second one the Bravais lattice type (Table 2). The number of atoms per unit cell is given at the third position. The symbol *cF*4, for instance, classifies a structure type to be cubic (*c*) all face centered (*F*) with four atoms per unit cell. In the case of rhombohedral structures, like the *hR*3-Sm type, the number of atoms in the unit cell of the rhombohedral setting ($a=b=c$, $\alpha=\beta=\gamma\neq 90^\circ$) is always given. The number of atoms in the corresponding hexagonal unit cell ($a=b\neq c$, $\alpha=\beta=90^\circ$, $\gamma=120^\circ$) is three times larger.

The atomic volume V_{at} is calculated by dividing the volume of the unit cell of a crystal structure by the number of atoms per unit cell. It is not only a measure of the actual size of an atom but also of the packing density q of a structure. The packing density always refers to a packing of hard spheres. In the case of just one size of spheres the maximum packing density is that of h.c.p. and c.c.p. structures and their polytypes (Fig. 1). The atomic volume V_{at} is just the volume of a spherical atom calculated from its radius divided by the packing density q. The very low value $q=\pi\sqrt{3}/16=0.34$ for the diamond structure, for

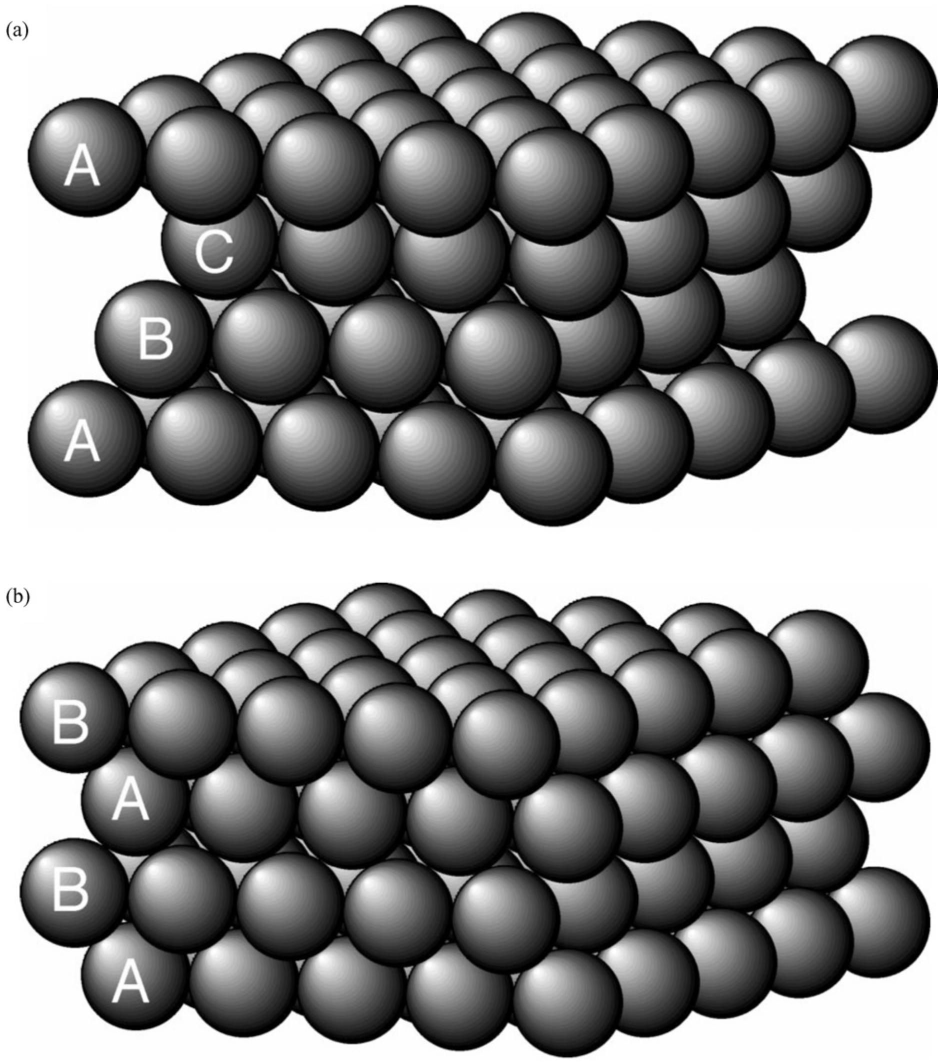


Figure 1
(a) Cubic and (b) hexagonal close sphere packings with packing densities of $q = 0.740$ in both cases. The packing sequence of the hexagonal close-packed layers is …ABC… in (a) and …AB… in (b). One body diagonal of the cubic unit cell is perpendicular to the h.c.p. layers in (a).

instance, does not reflect the high density of this material. It just demonstrates that the hard-sphere model is not suited properly to describe the packing of the nonspherical, covalently bonded carbon atoms.

2. *Discussion of the Crystal Structures of the Elements*

In the following, typical crystal structures of the elements are discussed by group. The crystallographic information of all allotropes of the elements being stable either at RT and arbitrary pressure or at RP and arbitrary temperature is compiled in Table 3. The data, however, are in a number of cases a somewhat subjective synthesis of different (sometimes contradictory) results but with emphasis on the most recent experimental and theoretical studies. Pressure-induced phase transformations, for instance, are often very sluggish and the experimentally observed transformation pressures may vary by several GPa.

2.1 Groups 1 and 2: Hydrogen, Alkali Metals, and Alkaline Earth Metals

In solid hydrogen, the h.c.p. H_2 molecules are rotating almost freely at low pressures. Under

Table 2
Meaning of the letters included in the Pearson notation.

Crystal family		Bravais lattice type	
a,	triclinic (anorthic)	*P*,	primitive
m,	monoclinic	*I*,	body centered
o,	orthorhombic	*F*,	all-face centered
t,	tetragonal	*S*,	side- or base-face centered
h,	hexagonal, trigonal (rhombohedral)	*R*,	rhombohedral
c,	cubic		

compression, rotational ordering transitions occur. A molecular metallic phase is expected at RT at pressures around 200 GPa and the onset of molecular dissociation around 620 GPa (Loubeyre *et al.* 1996).

The alkali metals are characterized by low melting temperatures and high compressibilities. The atomic volume of caesium, for instance, at 92 GPa amounts to just 21.4% of its value at RP. At low pressures, alkali metals behave like typical metals since they conform more closely to the free electron gas model of metals. The absence of directional bonds forces close-packed structures. At ambient conditions, the structures of alkali metals are all of the *cI*2-W type (Fig. 2). At low temperatures, lithium and sodium transform martensitically to phases with *hR*3-Sm type structures (Fig. 3).

At higher pressures, electron transfer from the half-filled $(n+1)s$ band to the empty *n*d band takes place for the heavier alkali metals (K, Rb, Cs, Fr). The *d* occupancy for caesium, for instance, was calculated to be 0.8 electrons per atom at 10 GPa (i.e., only 0.2 electrons per atom are left in the *s* band). The *s–d* crossover gives rise to the special softness of the heavy alkali metals and to a number of phase transitions.

The alkaline earth metals behave similarly to the alkali metals. One finds a comparable simple sequence of structural phase transitions for the light alkaline earth metals beryllium and magnesium, and a more complex behavior for the heavier ones. The large deviation of the hexagonal lattice parameter ratio $c/a = 1.56$ from the ideal value of $\sqrt{8/3} = 1.633$ indicates semicovalent bonding for Be(*hP*2)-I.

2.2 *Groups 3–10: Transition Metal (TM) Elements*

TM elements are typical metals with high melting temperatures and high hardness. Their partially occupied *nd* orbitals are only slightly screened by the outer $(n+1)s$ electrons. This leads to a significant change in the chemical properties of TMs going from left to right in the periodic system. The atomic volumes decrease rapidly with increasing number of electrons in bonding d orbitals due to increasing cohesion. They increase again, however, when the antibonding *d* orbitals are populated ("parabolic behavior"). At ambient conditions, TMs crystallize preferably in close-packed, high-symmetry structures. The exceptionally complex structures of the different manganese modifications result from its very narrow 3*d* bands. The bandwidths of the homologous 4*d* and 5*d* elements technetium and rhenium are considerably larger but still much narrower than the *sp* bands.

Group 3 elements scandium, yttrium, lanthanum, and actinium show similar phase sequences. High-pressure phases of the light elements occur as RP phases of the heavy homologues. La(*hP*4)-III, with stacking sequence …ACAB… of the h.c.p. layers, is one of the simpler polytypic close-packed structures common for the lanthanides (Fig. 3). Another typical polytype for lanthanides is Y(*hR*3)-III with stacking sequence …ABABCBCAC… (Fig. 3).

Group 4 elements titanium, zirconium, and hafnium all crystallize in slightly compressed h.c.p. structures and transform to b.c.c. at higher temperatures (RP). Under moderate compression, all three elements transform to phases with the open Ti(*hP*3)-III structure (ω-phase) (Fig. 4). Its packing density with $q = 0.57$ is slightly larger than that of Po(*cP*1)-II ($q = \pi/6 = 0.52$) but substantially lower than for b.c.c. ($q = \pi\sqrt{3}/8 = 0.68$) or c.c.p. and h.c.p. ($q = \pi/\sqrt{18} = 0.74$) structures. Calculations explain the stability of Ti(*hP*3)-III by covalent bonding contributions from 4*s*–3*d* electron transfer. At higher pressures, zirconium and hafnium transform to b.c.c., while titanium remains in the Ti(*hP*3)-III phase up to at least 87 GPa. The same type of h.c.p.–b.c.c. transformation that also occurs at high temperatures is of different origin in this case: the HT transformation is entropy driven while the HP transformation is caused by *s–d* electron transfer.

Group 5 and 6 elements vanadium, niobium, tantalum, molybdenum, and tungsten all have b.c.c. structures. Up to pressures of 170–364 GPa no further allotropes could be found in agreement with theoretical calculations. Chromium shows two antiferromagnetic phase transitions, which modify the structure only slightly.

Group 7 and 8 elements technetium, rhenium, ruthenium, and osmium all crystallize in h.c.p. structures. The HT phases Mn(*cF*4)-II and Mn(*cI*2)-I show typical metal structures while Mn(*cP*20)-III and Mn(*cI*58)-IV crystallize in very complex structures caused by their antiferromagnetism. Mn(*cI*58)-IV can be described as a $3 \times 3 \times 3$ superstructure of b.c.c. unit cells with 20 atoms slightly shifted and 4 atoms added resulting in total 58 atoms per unit cell.

Iron, the technologically most important element and main constituent of the earth's core, undergoes four phase transformations: ferromagnetic Fe(*cI*2)-III transforms to a paramagnetic isostructural phase; then Fe(*cF*4)-II forms, and below the melting point a

Table 3

Crystal structure types of the allotropes of the chemical elements. Only those HP and HT phases are listed that are stable at RT or RP, respectively. If no other reference is given, data are based on the compilations of Villars and Calvert (1991), Tonkov (1992, 1996), and Young (1991). Atomic volumes V_{at} are given at ambient conditions if no pressure p and temperature T is given in the form $T\vert p$. TP indicates triple point.

Element name, symbol, atomic number			Structure type, Pearson symbol, prototype formula			Stability range				
						Temperature (K)	Pressure (GPa)	V_{at} (Å^3)	$T\vert p$ (K)$\vert$(GPa)	Ref.[a]
Actinium	Ac	89	I	*cF*4	Cu	< 1324	RP	37.45		
Aluminum	Al	13	I	*cF*4	Cu	< 933.60	RP	16.61		
Americium	Am	95	γ, I	*cI*2	W	1350–1449	RP			
			β, II	*cF*4	Cu	1042–1350	RP	29.30		
						RT	5.2–12.5	25.69	RT $\vert$ 6.5	1
			α, III	*hP*4	La	< 1042	RP	29.27		
						RT	< 5.2			
			IV	*mP*4	Am	RT	12.5–15	24.00	RT $\vert$ 14.0	1
			δ, V	*oS*4	U	RT	> 15	23.35	RT $\vert$ 17.7	
Antimony	Sb	51	α, I	*hR*2	As	< 903.83	RP	30.21		
						RT	< 8			
			II	*tP*10	Sb	RT	8–28	24.47		
			III	*cI*2	W	RT	> 28			
Argon	Ar	18	I	*cF*4	Cu	< 83.798 TP	RP	37.45	4.25 $\vert$ RP	
Arsenic	As	33	α, I	*hR*2	As	< 887	RP	21.30		
						RT	< 25			
			II	*cP*1	Po	RT	25–48	16.58	RT $\vert$ 25	2
			III	*tP*10	Sb	RT	48–122	12.36	RT $\vert$ 106	2
			IV	*cI*2	W	RT	> 122			2
Astatine	At	85	I			< 575	RP			
Barium	Ba	56	α, I	*cI*2	W	< 1000	RP	62.99		
						RT	< 5.5			
			β, II	*hP*2	Mg	RT	5.5–7.5	40.55	RT $\vert$ 5.8	
			III	*cF*4	Cu	RT?	7.5–12.6			
			IV			RT	12.6–45			
			V	*hP*2	Mg	RT	> 45			
Berkelium	Bk	97	α, I	*hP*4	La	< 1250	RP	28.18		
						RT	< 8			
			β, II	*cF*4	Cu	1250–1323	RP			
						RT	7–25			
			III	*oS*4	U	RT	> 25			
Beryllium	Be	4	β, I	*cI*2	W	1523–1562	RP	8.30	1528 $\vert$ RP	
			α, II	*hP*2	Mg	< 1523	RP	8.11		
						RT	< 12			
			III	*oP*4	Be	RT	> 12	6.95	RT $\vert$ 28.3	
Bismuth	Bi	83	α, I	*hR*2	As	< 544.592	RP	35.07		3
						RT	< 2.55			3
			β, II	*mS*4	Bi	RT	2.55–2.69	31.62	RT $\vert$ 2.6	3
			γ, III	*tP*10	Bi	RT	2.69–7.7	30.61	RT $\vert$ 2.9	4
			ζ, IV	*cI*2	W	RT	> 7.7	27.44	RT $\vert$ 9	4
Boron	B	5	β, I	*hR*105	B	1453–2365	RP	23.67	298 $\vert$ RP	
			α, II	*hR*12	B	< 1453	RP	21.87		
						RT	< 210			
			III	*tI*		RT	210–360			
			IV	*cF*4	Cu	RT	> 360			
Bromine	Br	35	I	*oS*8	I	< 265.9 TP	RP	33.55	250 $\vert$ RP	
						RT	< 80			
			II	*oI*2	I	RT	> 80	15.10	RT $\vert$ 80	5

Table 3
Continued

Element name, symbol, atomic number			Structure type, Pearson symbol, prototype formula			Stability range		V_{at} (Å^3)	$T\mid p$ (K)|(GPa)	Ref.[a]
						Temperature (K)	Pressure (GPa)			
Cadmium	Cd	48	I	*hP*2	Mg	<594.258	RP	21.60		
Caesium	Cs	55	α, I	*cI*2	W	<301.54	RP	110.45		
						RT	<2.26			
			β, II	*cF*4	Cu	RT	2.26–4.15	53.57	RT|4.1	
			β′, III	*cF*4	Cu	RT	4.15–4.10	48.77	RT|4.17	
			γ, IV	*tI*4		RT	4.20–10	35.01	RT|8	
			V	*oS*16	Cs	RT	10–72	26.09	RT|25.8	6
			VI	*hP*2?		RT	>72	11.48	RT|92	7
Calcium	Ca	20	β, I	*cI*2	W	721–1115	RP	44.95	740|RP	
			α, II	*cF*4	Cu	<721	RP	43.63		
						RT	<19.5			
			III	*cP*1	Po	RT	32–42	17.88	RT|39	
			IV			RT	>42			
Californium	Cf	98	β, I	*cF*4	Cu	863–1173	RP	30.43		
			α, II	*hP*4	La	<863	RP	27.27		
						RT	<17			
			III	*aP*4	Cf	RT	30–41	17.12		
			IV	*oS*4	U	RT	>41	14.29	RT|46.6	
Carbon	C	6	I	*hP*4	C	<4100	RP	8.80		
						RT	<12			
			II	*cF*8	C	RT	12–1100	5.67		
Cerium	Ce	58	δ, I	*cI*2	W	999–1071	RP	34.97	1030|RP	
			γ, II	*cF*4	Cu	326–999	RP	34.37	300|RP	
			β, III	*hP*4	La	96–326	RP	34.78		
						RT	<5.1			
			α, IV	*cF*4	Cu	<96	RP	28.52	77|RP	
			α′, V	*oS*4	U	RT	5.1–12.2	23.84	RT|5.8	
			VI	*tI*2	In	RT	>12.2	12.75	RT|208	8
Chlorine	Cl	17	I	*oS*8	I	<172.18 TP	RP	28.86	113|RP	
Chromium	Cr	24	I	*cI*2	W	<2136	RP	12.00		
Cobalt	Co	27	α, I	*cF*4	Cu	695–1768	RP	11.34	723|RP	
			ε, II	*hP*2	Mg	<695	RP	11.07		
Copper	Cu	29	I	*cF*4	Cu	<1358.02		11.79		
Curium	Cm	96	β, I	*cF*4	Cu	1550–1618	RP	31.99		
						RT	23–43			
			α, II	*hP*4	La	<1550	RP	30.09		
						RT	<23			
			III	*oS*4	U	RT	>43	15.98	RT|45.5	
Dysprosium	Dy	66	β, I	*cI*2	W	1654–1685	RP	31.52		
			α, II	*hP*2	Mg	86–1654	RP	31.60		
						RT	<5			
			α′, III	*oS*4	U	<83.5	RP	31.55		
			γ, IV	*hR*3	Sm	RT	5–9	27.04	RT|7	
			V	*hP*4	La	RT	9–38			
			VI	*cF*4	Cu	RT	>38			
Einsteinium	Es	99	I	*cF*4	Cu	<1133				
Erbium	Er	68	I	*hP*2	Mg	<1802	RP	30.64		
						RT	<7			
			II	*hR*3	Sm	RT	7–13			
			III	*hp*4	La	RT	>13			

Table 3
Continued

Element name, symbol, atomic number			Structure type, Pearson symbol, prototype formula			Stability range				
						Temperature (K)	Pressure (GPa)	V_{at} (Å^3)	T|p (K)|(GPa)	Ref.[a]
Europium	Eu	63	I	*cI*2	W	<1095	RP	48.13		
						RT	<12.5	35.22	RT|5.9	
			II	*hP*2	Mg	RT	12.5–18	27.39	RT|15.1	
			III	*hP*36?		RT	>18	23.57	RT|20	
Fermium	Fm	100	I			<1800				
Fluorine	F	9	*β*, I	*cP*16	O	45.6–53.48 TP	RP	18.54	50|RP	
			α, II	*mS*8	F	<45.6	RP	16.05	23.1|RP	
Francium	Fr	87	I			<300				
Gadolinium	Gd	64	*β*, I	*cI*2	W	1508–1586	RP	33.22		
			α, II	*hP*2	Mg	<1508	RP	33.00		
						RT	<1.5			
			γ, III	*hR*3	Sm	RT	1.5–6.5	30.00	RT|3.5	
			IV	*hP*4	La	RT	6.5–24	27.68	573|7.5	
			V	*cF*4	Cu	RT	14–44			
			VI	*hP*6		RT	>44	15.19	RT|105.6	
Gallium	Ga	31	α, I	*oS*8	Ga	<302.9241 TP	RP	19.58	285.3|RP	
						RT	<3.0			
			II	*cI*12	Ga	RT	3.0–14	17.56	313|2.6	
			III	*tI*2	In	RT	14–120	15.00	RT|12	9
			IV	*cF*4	Cu	RT	>120	10.20	RT|115	9
Germanium	Ge	32	I	*cF*8	C	<1211.5	RP	22.64		
						RT	<9			
			II	*tI*4	Sn	RT	9–74	16.71	RT|12	
			III	*oI*4		RT	74–85	12.34	RT|81	10
			IV	*hP*4	La	RT	>85	10.24	RT|125	
Gold	Au	79	I	*cF*4	Cu	<1337.58	RP	16.96		
Hafnium	Hf	72	*β*, I	*cI*2	W	2013–2504	RP	23.62	2073|RP	
			α, II	*hP*2	Mg	<2013	RP	23.13		
						RT	<38			
			III	*hP*3	Ti	RT	38–71			
			IV	*cI*2	W	RT	>71			
Helium	^{4}He	2	α, II	*hP*2	Mg			28.45	3.96|0.013	
								5.68	300|23.3	
			β, I	*cF*4	Cu			19.06	16|0.125	
			γ, III	*cI*2	W			34.71	1.73|2.9	
Holmium	Ho	67	I	*cI*2	W	1701–1747	RP	31.05		
			α, II	*hP*2	Mg	<1701	RP	31.12		
						RT	<7.0			
			β, III	*hR*3	Sm	RT	7.0–13	25.87	RT|8.5	
			IV	*hP*4	La	RT	>13			
Hydrogen	H	1	I	*hP*4		<13.95	RP	19.02	4.5|RP	
Indium	In	49	I	*tI*2	In	<429.784	RP	26.18		
						RT	<45			
			II	*oF*4		RT	>45	15.00	RT|93	11
Iodine	I	53	I	*oS*8	I	<386.8	RP	42.63		
						RT	<21			
			II	*oI*2	I	RT	21–43	25.17	RT|30	
			III	*tI*4		RT	43–55	20.37	RT|49	
			IV	*cF*4		RT	>55	13.25	RT|276	12
Iridium	Ir	77	I	*cF*4	Cu	<2720	RP	14.14		

Table 3
Continued

Element name, symbol, atomic number			Structure type, Pearson symbol, prototype formula			Stability range				
						Temperature (K)	Pressure (GPa)	V_{at} (Å^3)	$T\|p$ (K)|(GPa)	Ref.[a]
Iron	Fe	26	δ, I	*cI*2	W	1667–1811	RP	12.60	1667|RP	
			γ, II	*cF*4	Cu	1185–1667	RP	12.13	1189|RP	
			α, III	*cI*2	W	<1185	RP	11.78		
						RT	<9			
			ε, IV	*hP*2	Mg	RT	>9	10.36	RT|15	
Krypton	Kr	36	I	*cF*4	Cu	<115.765	RP	46.81	92|RP	
Lanthanum	La	57	γ, I	*cI*2	W	1153–1191	RP	38.65	1160	
			β, II	*cF*4	Cu	533–1153	RP			
						RT	2.3–7	30.47	RT|6.5	
			α, III	*hP*4	La	<533	RP	37.42		
						RT	<2.3			
			IV	*hP*6		RT	7–60			
			V	*cF*4	Cu	RT	>60			13
Lawrencium	Lr	103	I			<1900				
Lead	Pb	82	α, I	*cF*4	Cu	<600.652	RP	30.33		
						RT	<13.2			
			β, II	*hP*2	Mg	RT	13.2–110	24.87	RT|13.9	
			III	*cI*2	W	RT	>110	14.40	RT|238	
Lithium	Li	3	I	*cI*2	W	70–453.8	RP	21.62		
			II	*hR*3	Sm	<70	RP	24.32	20|RP	
						RT	<6.9			
			III	*cF*4	Cu	RT	>6.9	14.83	RT|8	
Lutetium	Lu	71	I	*hP*2	Mg	<1936	RP	29.27		
						RT	<25			
			II	*hR*3	Sm	RT	25–45	21.13	RT|23	14
			III	*hP*4	La	RT	45–88			14
			IV	*hR*24		RT	>88	12.52	RT|142	14
Magnesium	Mg	12	I	*hP*2	Mg	<923	RP	23.24		
						RT	<50			
			II	*cI*2	W	RT	>50	12.88	RT|58	
Manganese	Mn	25	δ, I	*cI*2	W	1410–1519	RP	14.62	1413|RP	
			γ, II	*cF*4	Cu	1360–1410	RP	14.40	1368|RP	
			β, III	*cP*20	Mn	980–1360	RP	13.62	1008|RP	
			α, IV	*cI*58	Mn	<980	RP	12.21		
Mendelevium	Md	101	I			<1100				
Mercury	Hg	80	α, I	*hR*1	Hg	78–234.314	RP	23.07	83|RP	
			β, II	*tI*2	Pa	<78	RP	22.54	77|RP	
						RT	3.7–12			
			γ, III	*oP*4	Hg	RT	12–37	19.35	RT|19.8	15
			δ, IV	*hP*2	Mg	RT	>37	16.15	RT|40	15
Molybdenum	Mo	42	I	*cI*2	W	<2896	RP	15.58		
Neodymium	Nd	60	β, I	*cI*2	W	1128–1294	RP	35.22	1156|RP	
			α, II	*hP*4	La	<1128	RP	34.18		
						RT	<3.8			
			γ, III	*cF*4	Cu	RT	3.8–18	29.72	RT|5.5	
			IV	*hP*6	Pr	RT	18–39	24.21	RT|18	
			V	*mS*4		RT	>39			
Neon	Ne	10	I	*cF*4	Cu	<24.563 TP	RP	22.23	4.3|RP	
Neptunium	Np	93	γ, I	*cI*2	W	843–912	RP	21.81	873|RP	
						RT	<21			

Table 3
Continued

Element name, symbol, atomic number			Structure type, Pearson symbol, prototype formula			Stability range				
						Temperature (K)	Pressure (GPa)	V_{at} (Å^3)	$T\|p$ (K)|(GPa)	Ref.[a]
			II	*oI*2	I	RT	21–43	25.17	RT|30	
			III	*tI*4		RT	43–55	20.37	RT|49	
			IV	*cF*4		RT	>55	13.25	RT|276	12
Iridium	Ir	77	I	*cF*4	Cu	<2720	RP	14.14		
Iron	Fe	26	δ, I	*cI*2	W	1667–1811	RP	12.60	1667|RP	
			γ, II	*cF*4	Cu	1185–1667	RP	12.13	1189|RP	
			α, III	*cI*2	W	<1185	RP	11.78		
						RT	<9			
			ε, IV	*hP*2	Mg	RT	>9	10.36	RT|15	
Osmium	Os	76	I	*hP*2	Mg	<3306	RP	13.98		
Oxygen	O	8	γ, I	*cP*16	O	43.8–54.4	RP	19.91	50|RP	
			β, II	*hR*2	O	23.85–43.8	RP	17.76		
						RT	5.4–9.3	11.96	RT|5.5	
			α, III	*mS*4	O	<23.85	RP	17.31	23|RP	
			δ, IV	*oF*8		RT	9.3–9.8	10.42	RT|9.6	
			ε, VI	*mP*4		RT	>9.8	8.83	RT|10	
Palladium	Pd	46	I	*cF*4	Cu	<1828	RP	14.69		
Phosphorus	P	15	I	*oS*8	P	<862.8	RP	18.99		
						RT	<5.5			
			II	*hR*2	As	RT	5.5–10	14.33	RT|9	
			III	*cP*1	Po	RT	>10	12.00	RT|25	
Platinum	Pt	78	I	*cF*4	Cu	<2042.2	RP	15.10		
Plutonium	Pu	94	ε, I	*cI*2	W	756–913	RP	24.07	773|RP	
			δ′, II	*tI*2	In	725–756	RP	24.78	750|RP	
			δ, III	*cF*4	Cu	583–725	RP	24.89	653|RP	
			γ, IV	*oF*8	Pu	488–583	RP	23.14	563|RP	
			β, V	*mS*34	Pu	388–488	RP	22.16	366|RP	
			α, VI	*mP*16	Pu	<388	RP	20.43		
						RT	<40			
			VII	*hP*8		RT	>40	13.95	RT|62	17
Polonium	Po	84	β, I	*hR*1	Po	327–527	RP	36.89		
			α, II	*cP*1	Po	<327	RP	37.90		
Potassium	K	19	I	*cI*2	W	<336.86	RP	75.72	301|RP	
						RT	<11.4			
			II	*cF*4	Cu	RT	11.4–23	34.11	RT|12.4	
			III	*tI*16		RT	>23	25.98	RT|29.3	18
Praseodymium	Pr	59	β, I	*cI*2	W	1069–1204	RP	35.22	1094|RP	
			α, II	*hP*4	La	<1069	RP	34.52		
						RT	<3.8			
			γ, III	*cF*4	Cu	RT	3.8–6.2	29.05	RT|4	
			IV	*hR*8		RT	6.2–14	24.32	RT|13.8	19
			V	*oS*4	U	RT	>14	18.91	RT|26.2	
Promethium	Pm	61	β, I	*cI*2	W	1163–1315	RP			
			α, II	*hP*4	La	<1163	RP	33.85		
						RT	<10			
			III	*cF*4	Cu	RT	10–18			
			IV	*hP*6	Pr	RT	18–40			
			V			RT	>40			
Protactinium	Pa	91	β, I	*cI*2	W	1443–1845	RP	27.65		
			α, II	*tI*2	Pa	<1443	RP	25.64		

Table 3
Continued

Element name, symbol, atomic number			Structure type, Pearson symbol, prototype formula			Stability range		V_{at} (Å^3)	$T\|p$ (K)|(GPa)	Ref.[a]
						Temperature (K)	Pressure (GPa)			
Radium	Ra	88	I	*cI*2	W	<973	RP	68.22		
Radon	Rn	86	I			<202	RP			
Rhenium	Re	75	I	*hP*2	Mg	<3459	RP	14.71		
Rhodium	Rh	45	I	*cF*4	Cu	<2236	RP	13.77		
Rubidium	Rb	37	α, I	*cI*2	W	<312.63	RP	87.10		
						RT	<7.0			
			II	*cF*4	Cu	RT	7.0–13			
			III			RT	13–16.5			
			IV			RT	16.5–19.5			
			V	*tI*4		RT	19.5–46			
			VI			RT	>46			18
Ruthenium	Ru	44	I	*hP*2	Mg	<2607	RP	13.57		
Samarium	Sm	62	γ, I	*cI*2	W	1195–1347	RP			
			β, II	*hP*2	Mg	1007–1195	RP	33.79	980|RP	
			α, III	*hR*3	Sm	<1007	RP	33.18		
						RT	<4.5			
			δ, IV	*hP*4	La	RT	4.5–14	33.05		
			V	*cF*4	Cu	RT	14–19			
			VI	*hR*8	Pr	RT	19–33			
			VII	*mS*4		RT	33–91			
			VIII	*tI*2		RT	>91	12.21	RT|189	20
Scandium	Sc	21	β, I	*cI*2	W	1609–1814	RP	25.95		
			α, II	*hP*2	Mg	<1609	RP	25.00		
						RT	<19			
			III	*tP*4	Np	RT	>19	16.81	RT|26	
Selenium	Se	34	γ, I	*hP*3	Se	<494	RP	27.28		
						RT	<14	21.41	RT|7.7	
			II			RT	14–23			21
			III	*mS*4		RT	23–28	17.37	RT|23	21
			IV	*oS*4		RT	28–60	15.77	RT|34.9	21
			V	*hR*1	Po	RT	60–140	14.35	RT|60	21
			VI	*cI*2	W	RT	>140	11.25	RT|140	21
Silicon	Si	14	α, I	*cF*8	C	<1687	RP	20.02		
						RT	<12			22
			β, II	*tI*4	Sn	RT	12–13.2	14.19	RT|13	22
			III	*oI*4		RT	13.2–15.6	13.60	RT|15	23
			IV	*hP*1	BiIn	RT	15.6–37.6	12.28	RT|34	23
			V	*oS*16	Cs	RT	37.6–42	11.49	RT|38.4	23
			VI	*hP*2	Mg	RT	42–78	10.74	RT|43	23
			VII	*cF*4	Cu	RT	>78	9.27	RT|87	23
Silver	Ag	47	I	*cF*4	Cu	<1235.08	RP	17.06		
Sodium	Na	11	β, I	*cI*2	W	36–371.0	RP	39.49		
			α, II	*hR*3	Sm	<36	RP	43.57	20|RP	
Strontium	Sr	38	β, I	*cI*2	W	896–1042	RP	57.75	901|RP	
						RT	3.5–26			
			II	*hP*2	Mg	504–896	RP	57.05	521|RP	
			α, III	*cF*4	Cu	<504	RP	56.08		
						RT	<3.5			
			IV	*oI*4		RT	26–35	24.53	RT|31.3	24
			V			RT	35–46			

Table 3
Continued

Element name, symbol, atomic number			Structure type, Pearson symbol, prototype formula			Stability range		V_{at} (Å^3)	$T\|p$ (K)|(GPa)	Ref.[a]
						Temperature (K)	Pressure (GPa)			
			VI			RT	>46			
Sulfur	S	16	β, I	*mP*48	S	368.65–388.37	RP	27.23	376|RP	
			α, II	*oF*128	S	<368.65	RP	25.76		
						RT	<5.3			
			III	*mP*		RT	5.3–24			25
			IV	amorphous		RT	24–37			25
			V				37–84			25
			VI	*oS*4		RT	84–162	8.82	RT|145	25
			VII	*hR*1	Po	RT	>162	8.01	RT|206.5	25
Tantalum	Ta	73	I	*cI*2	W	<3293	RP	18.01		
Technetium	Tc	43	I	*hP*2	Mg	<2428	RP	14.31		
Tellurium	Te	52	α, I	*hP*3	Se	<722.72	RP	34.00		
						RT	<4.2	31.30	RT|2	26
			II	*mP*4		RT	4.2–6.8	28.00	RT|5	26
			III	*oP*4		RT	6.8–10.6	26.20	RT|8.5	26
			IV	*hR*1	Po	RT	10.6–27	23.60	RT|17.5	26
			VII	*cI*2	W	RT	>27	20.65	RT|33.00	26
Terbium	Tb	65	β, I	*cI*2	W	1562–1629	RP	32.48		
			α′, II	*hP*2	Mg	220–1562	RP	32.13		
						RT	<3.0			
			α, III	*oS*4	Dy	<220	RP	32.11	223|RP	
			γ, IV	*hR*3	Sm	RT	3.0–6.0	27.41	RT|6.0	
			V	*hP*4	La	RT	6.0–29			
			VI	*cF*4	Cu	RT	29–32			
			VII	*hR*8	Pr	RT	>32			
Thallium	Tl	81	β, I	*cI*2	W	505–577	RP	29.00	523|RP	
			α, II	*hP*2	Mg	<503	RP	28.60		
						RT	<3.7			
			γ, III	*cF*4	Cu	RT	>3.7	27.27	RT|6	
Thorium	Th	90	β, I	*cI*2	W	1633–2028	RP	34.71	1723|RP	
			α, II	*cF*4	Cu	<1633	RP	32.89		
						RT	<100			
			III	*tI*2	In	RT	>100	13.32	RT|300	27
Thulium	Tm	69	I	*cI*2	W	1800–1818	RP	30.12		
			II	*hP*2	Mg	<1800	RP	30.27		
						RT	<9			
			III	*hR*3	Sm	RT	9–13			
			IV	*hP*4	La	RT	>13	25.01	RT|11.6	
Tin	Sn	50	β, I	*tI*4	Sn	291–505.1181	RP	27.05		
						RT	<9.4			
			α, II	*cF*8	C	<291	RP	34.16	293|RP	
			γ, III	*tI*2	Pa	RT	9.4–45	20.25	RT|24.5	
			IV	*cI*2	W	RT	>45	17.76	RT|53	
Titanium	Ti	22	β, I	*cI*2	W	1155–1943	RP	18.07	1193|RP	
			α, II	*hP*2	Mg	<1155	RP	17.64		
						RT	<2			
			ω, III	*hP*3	Ti	RT	>2	17.23	RT|4	
Tungsten	W	74	I	*cI*2	W	<3695	RP	15.85		
Uranium	U	92	γ, I	*cI*2	W	1049–1408	RP	22.22	1173|RP	
			β, II	*tP*30	CrFe	941–1049	RP	21.81	955|RP	

Table 3
Continued

Element name, symbol, atomic number			Structure type, Pearson symbol, prototype formula			Stability range		V_{at} (Å^3)	$T \mid p$ (K)|(GPa)	Ref.[a]
						Temperature (K)	Pressure (GPa)			
			α, III	*oS*4	U	<941	RP	20.75		
Vanadium	V	23	I	*cI*2	W	<2183	RP	13.87		
Xenon	Xe	54	I	*cF*4	Cu	<161.3918 TP	RP	59.50	58|RP	
						RT	<21			
			II	*hP*2	Mg	RT	>21	17.35	RT|137	28
Ytterbium	Yb	70	γ, I	*cI*2	W	1047–1092	RP	43.76	1047|RP	
						RT	3.5–26			
			β, II	*cF*4	Cu	270–1047	RP	41.21		
						RT	<3.5			
			α, III	*hP*2	Mg	<270	RP	41.63		
						RT	26–53	18.98	RT|34	
			IV	*cF*4		RT	53–98	16.36	RT|57	29
			V	*hP*3	Sm	RT	>98	10.85	RT|202	30
Yttrium	Y	39	β, I	*cI*2	W	1751–1795	RP	34.71		
			α, II	*hP*2	Mg	<1751	RP	33.18		
						RT	<10			
			III	*hR*3	Sm	RT	10–26			
			IV	*hP*4		RT	26–39			
			V	*cF*4	Cu	RT	>39			
Zinc	Zn	30	I	*hP*2	Mg	<692.73	RP	15.21		
Zirconium	Zr	40	β, I	*cI*2	W	1136–2128	RP	23.50		
			α, II	*hP*2	Mg	<1136	RP	23.30		
						RT	<2.2			
			ω, III	*hP*3	Ti	RT	2.2–30			
			IV	*cI*2	W	RT	>30			

[a] 1, Roof (1981); 2, Iwasaki (1997); 3, Iwasaki and Kikegawa (1997); 4, Chen *et al.* (1996); 5, Fujihisa *et al.* (1995); 6, Schwarz *et al.* (1998); 7, Takemura *et al.* (1991); 8, Vohra *et al.* (1999); 9, Takemura *et al.* (1998); 10, Nelmes *et al.* (1996); 11, Takemura and Fujihisa (1993); 12, Reichlin *et al.* (1994); 13, Porsch and Holzapfel (1993); 14, Chesnut and Vohra (1998); 15, Schulte and Holzapfel (1993); 16, Olijnyk (1990); 17, Dabos-Seignon *et al.* (1993); 18, Winzenick *et al.* (1994); 19. Hamaya *et al.* (1993); 20, Vohra and Akella (1991); 21, Akahama *et al.* (1993); 22, Hanfland *et al.* (1999); 23, McMahon and Nelmes (1993); 24, Winzenick and Holzapfel (1996); 25, Luo *et al.* (1993); 26, Parthasarathy and Holzapfel (1988); 27, Vohra *et al.* (1991); 28, Caldwell *et al.* (1997); 29, Zhao *et al.* (1994); 30, Chesnut and Vohra (1999).

b.c.c. phase, Fe(*cI*2)-I, appears again; Fe(*cI*2)-III transforms under compression into Fe(*hP*2)-IV. A further orthorhombic phase has been found to be stable at high pressures and temperatures (HTP).

Group 9 and 10 elements rhodium, iridium, nickel, palladium, and platinum all show f.c.c. structures. Cobalt undergoes a phase transformation from h.c.p. (RT) to c.c.p. at HT. By annealing it in a special way, stacking disorder can be generated: the h.c.p. sequence …AB… is statistically replaced by a c.c.p. sequence …ABC… such as …ABABABABCBCBCBC… with a frequency of about one …ABC… among ten …AB….

2.3 Groups 11 and 12: Copper and Zinc Group Metals

The coinage metals copper, silver, and gold are typical f.c.c. metals (Fig. 5). Their single *ns* electron is less shielded by filled $(n-1)d$ orbitals than the *s* electron of the alkali metals by the fully occupied noble gas shells. The *d* electrons also contribute to the metallic bond. These factors are responsible for the more noble character of these metals compared to the alkali metals.

Zinc and cadmium do not undergo phase transitions up to the highest pressures employed (Zn: 126 GPa, Cd: 174 GPa). The c/a ratios, however, decrease from 1.856 for zinc and 1.886 for cadmium to 1.59. The value for an ideal h.c.p. structure is $\sqrt{8/3} = 1.633$, and most of the h.c.p. metallic elements fall in the range $1.57 < c/a < 1.65$. The large deviation from the ideal value at RTP is due to a lowering of the band structure energy by lattice distortions (Peierl's distortion).

The rhombohedral structure of Hg(*hR*1)-I can be derived from a f.c.c. structure by strong compression along the threefold axis (Fig. 3). In contrast to zinc

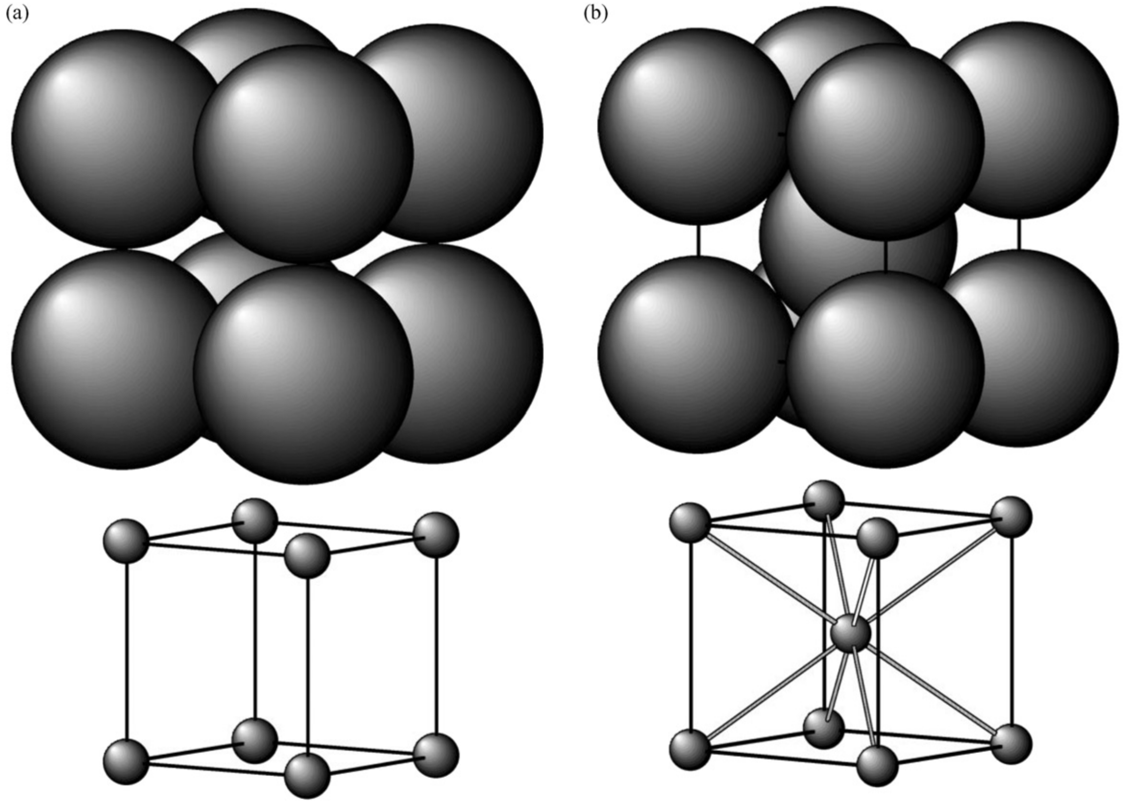


Figure 2
Cubic structures with packing densities of (a) 0.523 for the *cP*1-Po type and (b) 0.680 for the *cI*2-W type. By "full-size" (100%) atoms the interatomic contacts can be better visualized (upper), while the "stick-and-ball" drawings (20%) are better suited to illustrate more complex structures (lower).

and cadmium, the ratio $c/a+=1.457$ for a description as a hypothetical distorted h.c.p. structure is smaller than the ideal value.

2.4 Group 13: Boron Group Elements

There is still not very much reliable information on the phase diagram of boron. Its structural chemistry, dominated by icosahedral structural units, is complicated and similar to that of sulfur. Aluminum and thallium, however, are simple *sp*-bonded metals and crystallize in close-packed structures. Ga(*oS*8)-I forms a 6^3 network of distorted hexagons. The bonds between these hexagon layers are considerably weaker than within the layers. At higher pressures, gallium transforms to phases with b.c.c. structures and finally to Ga(*cF*4)-V.

2.5 Group 14: Carbon Group Elements

Graphite, hexagonal C(*hP*4)-I, is the stable carbon phase at RTP (Fig. 4). Under strong compression C(*cF*8)-II, diamond (Fig. 5), is formed at RT. The exceptionally hard diamond structure is denser than the predicted zero-pressure states of the close-packed metals due to the strongly *p*-character sp^3 bonding electrons (Young 1991).

Semiconducting silicon, germanium, and the low-temperature phase of tin crystallize in the diamond structure because of strong covalent bonds with tetrahedral sp^3 hybridization. Under compression, they transform to metallic phases with the white-tin (*tI*4-Sn) structure (Fig. 6). This structural type can be regarded as being intermediate between the diamond structure of Sn(*cF*8)-II and f.c.c. Pb(*cF*4)-I. For an ideal ratio of $c/a=0.528$ one atom is sixfold coordinated. The high-pressure phase Si(*hP*1)-IV has a quasi-eightfold coordination with $c/a=0.92$. The ideal ratio for eightfold coordination would be $c/a=1$. At higher pressures close-packed structures with 12-fold coordination are obtained. Thus, with increasing pressure, tin atoms run through coordination numbers 4, 6, 8, and 12.

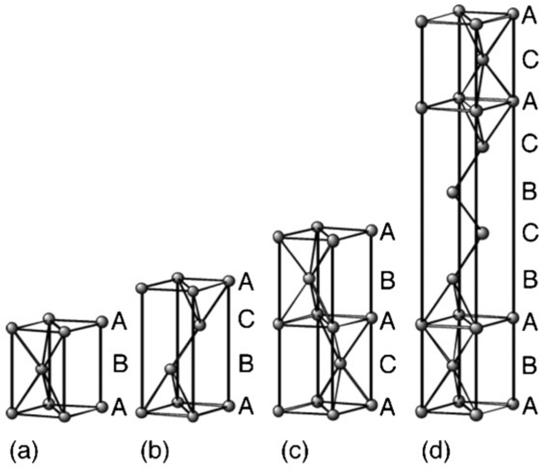


Figure 3
Polytypic structure types: (a) *hP*2-Mg, (b) *hR*1-Hg, (c) *hP*4-La, (d) *hR*3-Sm. Packing sequences of the hexagonal close-packed layers are indicated. Sizes of spheres correspond to 20% atomic diameter.

The differences in the structural behavior of carbon with respect to silicon and germanium increase with pressure. This is due to d electron contributions in silicon and germanium. The different behavior of lead can be explained by relativistic effects (Holzapfel 1996). The effective radii of tin in Sn(*tI*4)-I and of lead in Pb(*cF*4)-I are large compared to that of other typical metals with large atomic number. This is due to incomplete ionization of the single *s* electron. This means that in Sn(*cF*8)-II, for instance, the electron configuration is $\ldots 5s^1 5p^3$, allowing sp^3 hybridization and covalent tetrahedrally coordinated bonding. In Sn(*tI*4)-I, with electron configuration $\ldots 5s^2 5p^2$, however, only two *p* orbitals are available for covalent and one further *p* orbital for metallic bonding.

2.6 *Group 15: Nitrogen Group Elements*

Solid N_2 shows a very rich phase diagram with two low-temperature (LT) phases at RP, ordered N(*cP*8)-II and orientationally disordered N(*hP*4)-I. There are also several HP phases known. Theoretical calculations give some evidence for a similar sequence of phases, from the *hR*2-As to the *cP*1-Po structure type, as found for the heavier homologues.

Black phosphorus, P(oS8)-I, is the stable phosphorus phase at RTP while the red modification is the HT phase (Brazhkin and Zerr 1992). The hR2-As-type structures of arsenic, antimony, and bismuth consist of puckered layers of covalently bonded atoms stacked along the hexagonal axis (Fig. 7). The structures can be regarded as distorted cP1-Po-type

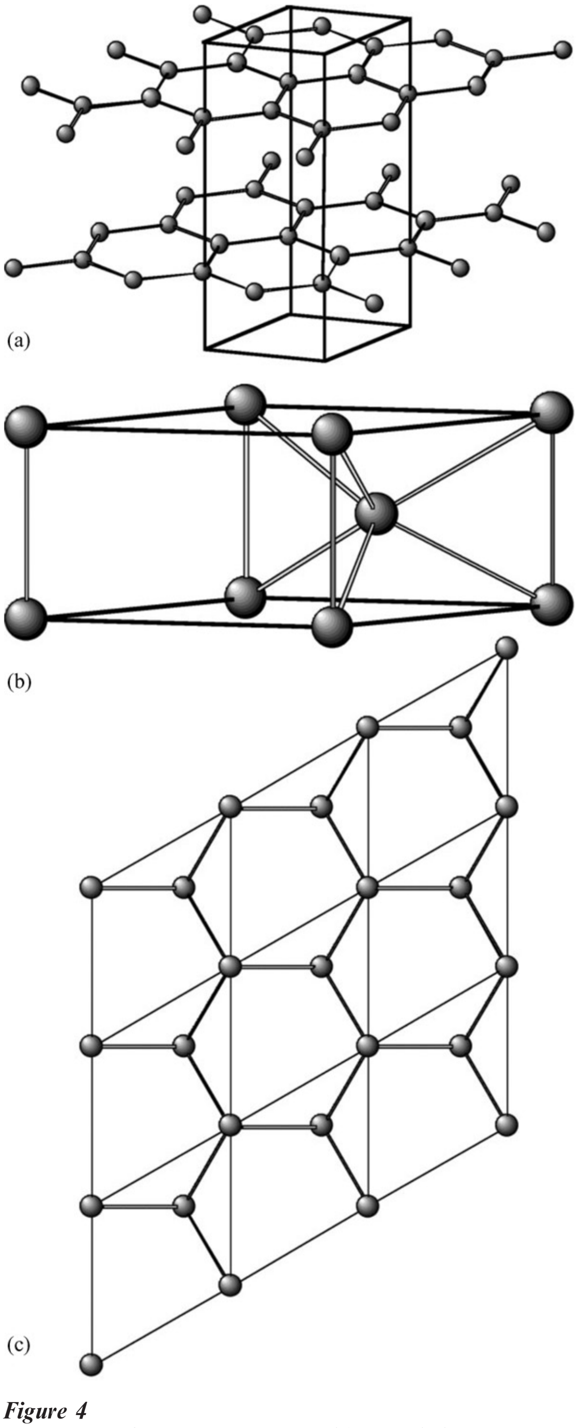


Figure 4
Structures of (a) C(*hP*4)-I (graphite) and (b), (c) Ti(*hP*3)-III (*ω*-titanium). The carbon atoms form planar 6^3-nets, the titanium atoms staggered ones. In (c) the *hP*3-Ti structure is shown projected down *c*. Sizes of spheres correspond to 20% atomic diameter.

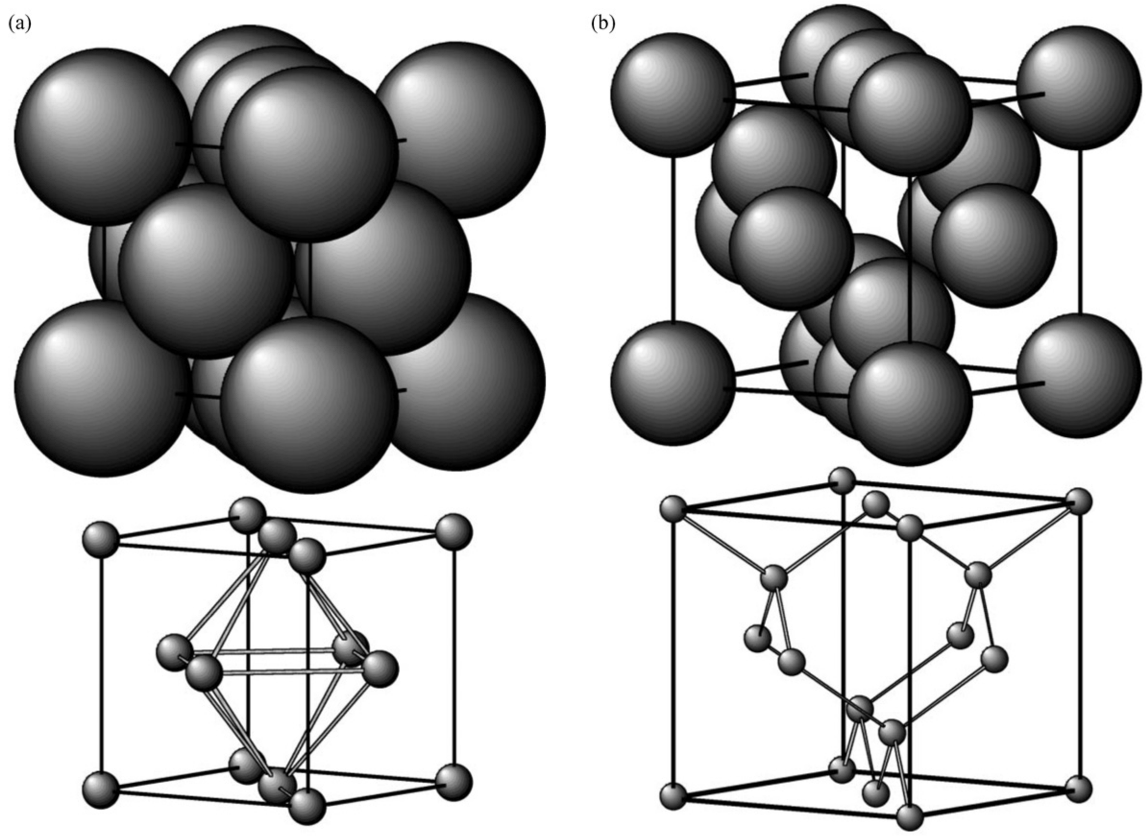


Figure 5
Cubic sphere packings with packing densities (a) $q = 0.740$ for the close-packed *cF*4-Cu type and (b) $q = 0.340$ for the *cF*8-C (diamond) structure type. The actual density of the diamond structure is much higher because of the nonspherical shape of carbon atoms (*sp*^{3} bonding). Sizes of spheres correspond to 100% atomic diameter (upper) and to 20% (lower).

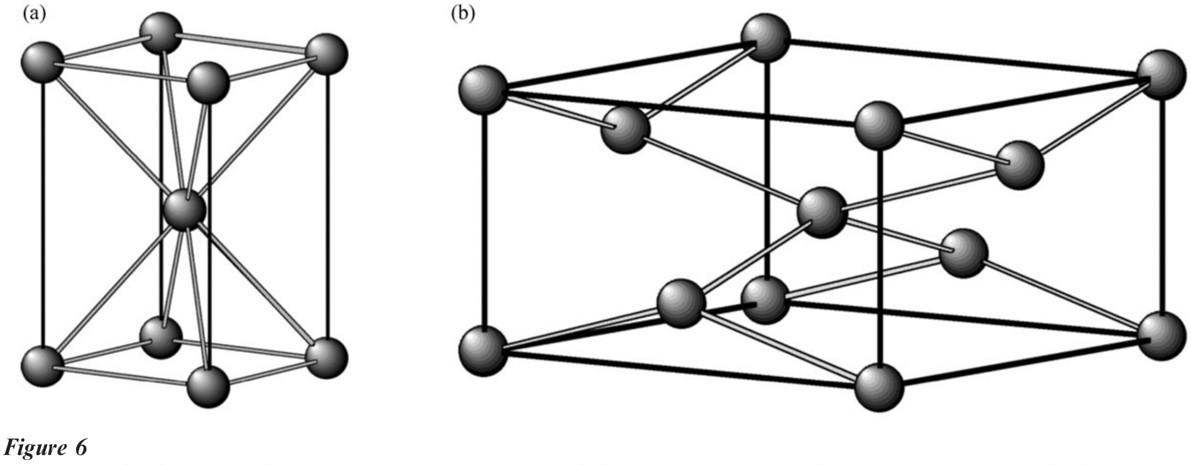


Figure 6
Tetragonal body-centered structure types: (a) *tI*2-In and (b) *tI*4-Sn (metallic white tin, β-tin). Sizes of spheres correspond to 20% atomic diameter.

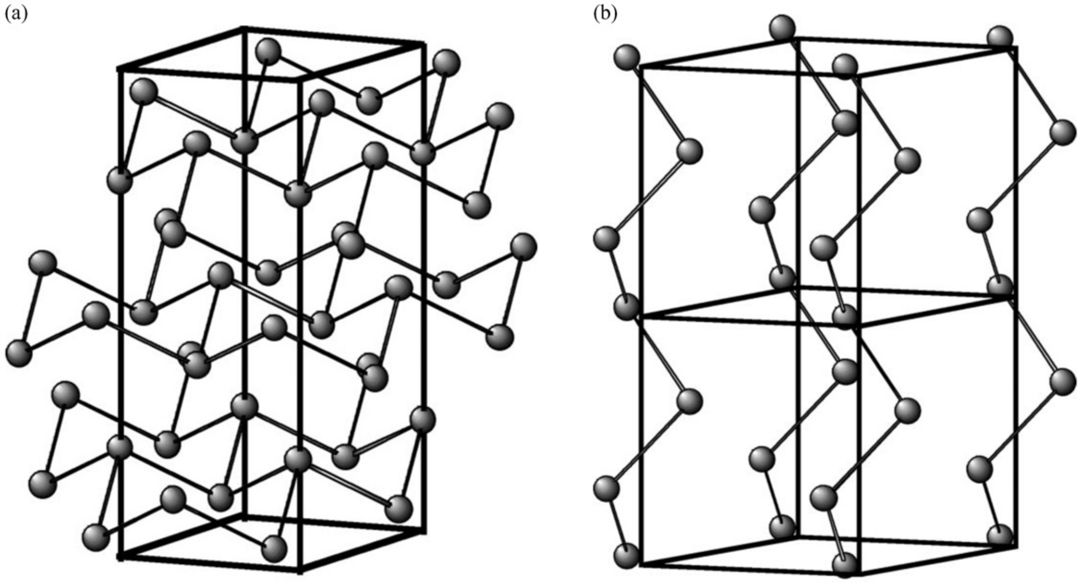


Figure 7
Structures of (a) As($hR2$)-I (gray, metallic α-arsenic) and (b) Se($hP3$)-I (metallic α-selenium). Sizes of spheres correspond to 20% atomic diameter.

structures in which the atomic distance d_1 in the layers equals d_2, the distance between the layers. The metallic character of these elements increases as d_2/d_1 approaches unity. Strong relativistic effects appear to be responsible for the complexity of the bismuth phase diagram (Holzapfel 1996).

2.7 Group 16: Chalkogens

With increasing atomic number the chalkogens shift from diatomic molecular solids (O_2), to polymeric molecular solids (S_n, Se_n), then to semimetallic solids (tellurium), and finally to metallic solids (polonium). Molecular solid oxygen and, especially, sulfur show extremely rich and complicated phase diagrams. In the case of oxygen, magnetic ordering plays a structure-stabilizing role: O($mS4$)-III shows antiferromagnetic ordering and O($hR2$)-II quasihelical magnetic ordering.

The trigonal helical structures of Se($hP3$)-I and Te($hP3$)-I can be derived from the Po($cP1$) structure (Figs. 2 and 7). Infinite helices run along the trigonal axes, and have three atoms per turn. The interhelix bonding distance is a measure of the metallic character of these elements. The semiconductor–metal transitions (Se: 20 GPa, Te: 4 GPa) are of strongly first-order nature.

The open structure of Po($cP1$)-II (Fig. 2) is stabilized by relativistic spin–orbit effects (splitting of the 6*s* band and suppression of covalent bonds) (Holzapfel 1996). The structure of the high-temperature phase Po($hR1$)-I represents just a slight elongation of the cubic structure along one of its threefold axes.

2.8 Group 17: Halogens

Solid halogens show rather close-packed molecular structures at RP. Under compression, the diatomic molecules dissociate (Cl_2: ~220 GPa, Br_2: ~80 GPa, I_2: ~20 GPa) and higher-symmetry monatomic structures are formed. Metallic I($cF4$)-IV, for instance, is stable at least up to 276 GPa and superconducting at low temperatures. Semiconductor–metal transitions, such as that observed around 16–18 GPa for I_2, are expected around 100 GPa for Br_2. In the case of I_2, metallization occurs by closure of the 5*p*–5*d* band gap even before dissociation of molecules takes place (Fig. 8). Theoretical calculations indicate a considerable ns–$(n-1)d$ electron transfer under compression for Cl_2, Br_2, and I_2 (Fujihisa *et al.* 1995). The high-pressure behavior of F_2 is expected to differ distinctly from that of the other halogens with their *s*–*d* electron transfer under strong compression.

While the geometrical electronegativities of metals increase monotonically with pressure, those of halogens show a sharp decrease until the molecules dissociate. Subsequently, the electronegativity increases slowly again. This can be explained qualitatively by the simple assumption that the decrease of the atomic volume under pressure causes an increase of the effective nuclear charge. This is followed by the increase of the covalent radius of the atoms after the dissociation of the molecules (Batsanov 1997).

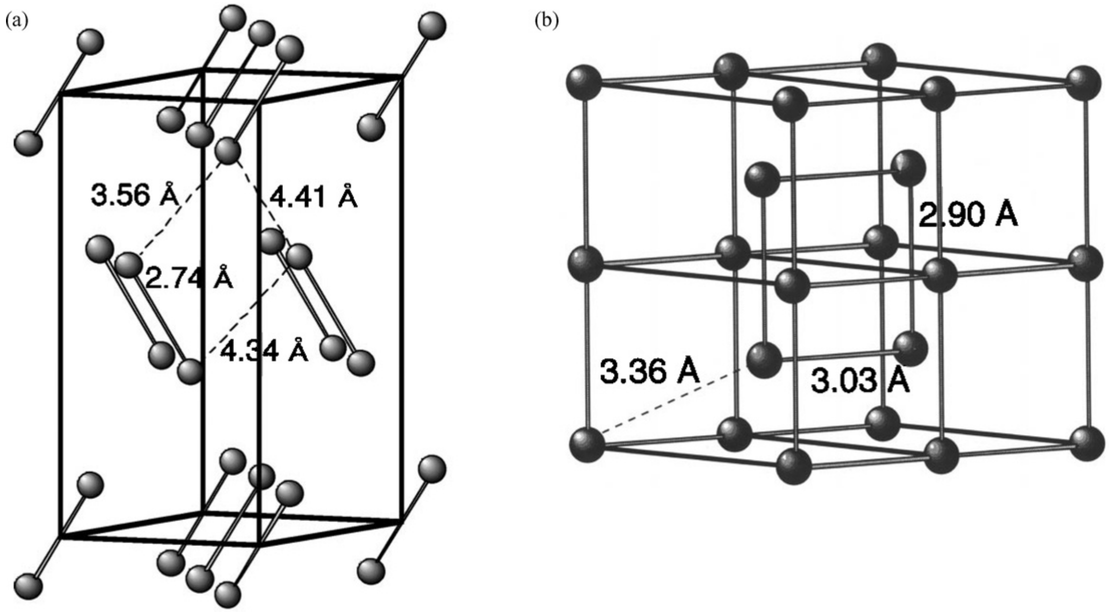


Figure 8
Crystal structures of the orthorhombic diatomic low-pressure and the monatomic high-pressure phase of iodine. (a) Side-centered I(*oS*8) (molecular iodine) and (b) body-centered I(*oI*2)-II (metallic iodine). The strongly varying distances marked in (a) become almost equal in (b). Sizes of spheres correspond to 20% atomic diameter.

2.9 Group 18: Noble Gases

Helium is the only element that does not become solid at RP. The reason is large-amplitude, zero-point vibrations due to the weak interactions between helium atoms together with their small mass. Under compression, the atomic interactions become more complex and structural phase transitions occur (He, Xe) or are calculated to occur at very high pressures (Ne, Ar, Kr, Rn). All structures are h.c.p., c.c.p., or b.c.c.. Finally, at very high pressures the noble gases are expected to become metallic. In the case of Xe(*hP*2)-II, metallization has already been observed around 150–160 GPa. It can be explained by closure of the 5*p*–5*d* band gap.

2.10 Lanthanides (Rare Earth (RE) Elements)

Lanthanides and actinides are characterized by the fact that their *nf* valence electrons are screened by filled $(n+1)s$ and $(n+1)p$ orbitals. This leads to rather uniform chemical properties. The chemical behavior of the actinides, however, is somewhat between that of the 3*d* transition metals and the lanthanides since the 5*f* orbitals are screened to a much smaller amount by the 6*s* and 6*p* electrons. By gradually filling the 4*f* orbitals the atomic volumes slowly decrease, mainly due to the incomplete shielding of the nuclear charge by the 4*f* electrons (“lanthanide contraction”). A similar shrinkage of the atomic volumes is observed when the 5*f* orbitals are filled across the actinide series (“actinide contraction”).

With the exception of samarium and europium, all lanthanides show, at ambient conditions, either simple h.c.p. structures with the standard stacking sequence …AB… or twofold superstructures (d.h.c.p.) with stacking sequence …ACAB…. Sm-III has, with …ABABCBCAC…, an even more complicated stacking sequence with a 4.5-fold superperiod (Fig. 3). For all lanthanides, the ratio c/a is close to the ideal value of $n \times 1.633$. With increasing pressure and decreasing atomic number the sequence of close-packed phases with structure types (polytypes) *hP*2-Mg (…AB…)→ *hR*3-Sm (…ABABCBCAC…)→*hP*4-La (…ACAB…)→*cF*4-Cu (…ABC…)→*hP*6-Pr appears in all regular “trivalent” RE metals (Fig. 3). It is dominated by 5*s*–5*d* electron transfer (and only minor 4f electron contributions). Scandium, divalent europium, ytterbium, and irregular (due to *f* band delocalization) cerium show different structural behavior.

Cerium has an extraordinarily rich phase diagram. It is the first metal with an occupied 4*f* orbital. Its valence electrons can either be localized as in Ce(*cF*4)-II or itinerant as in Ce(*cF*4)-IV. There is an isostructural transition with 15% volume collapse between the two f.c.c. phases.

Europium shows a behavior different from the other lanthanides due to the stability of its half-filled

4*f* orbitals. This makes it more similar to the alkaline earth metals. The phase diagram of europium is comparable to that of barium rather than to those of the other RE elements. A similar behavior is observed for ytterbium which is divalent due to the stability of the completely filled 4*f* orbitals, and whose phase diagram resembles that of strontium.

2.11 Actinides

According to their electronic properties, the actinides can be divided into two subgroups: the light actinides, from thorium to plutonium, have itinerant 5*f* electrons contributing to the metallic bonding; the heavier actinides, from americium onwards have more localized 5*f* electrons. Under moderate pressures, the heavier actinides change from regular RE metal behavior with localized 5*f* electrons to cerium-type behavior with delocalized 5*f* electrons. This situation leads to superconductivity for thorium, protactinium, and uranium, and to magnetic ordering for curium, berkelium, and californium (Dabos-Seignon *et al.* 1993). The contribution of 5*f* electrons to bonding leads to low-symmetry structures, small atomic volumes, and high density in the case of the light actinides while the heavier actinides crystallize at ambient conditions in the h.c.p. structure. The position of plutonium at the border of itinerant and localized 5*f* states causes its unusually complex phase diagram with structures typical for both cases. Thus monoclinic Pu(*mP*16)-VI, for instance, can be considered as a distorted h.c.p. structure with about 20% higher packing density than f.c.c. Pu(*cF*4)-III. This is the result of covalent bonding contributions from 5*f* electrons.

The phase diagram of americium is very similar to those of lanthanum, praseodymium, and neodymium. Americium is the first lanthanide-like actinide element because of the localization of 5*f* electrons. There is a volume collapse under pressure for curium, berkelium, and californium, but not for americium (not yet understood theoretically) (Johansson 1995).

Thorium behaves similarly to cerium. Under compression, electron transfer from the 6*spd* band to the 5*f* band takes place. This is responsible for the continuous transition from Th(*cF*4)-II to tetragonal Th(*tI*2)-III. The corresponding transition occurs in cerium at lower pressures due to its higher 4*f* band population (Söderlind 1998).

Bibliography

Akahama Y, Kobayashi M, Kawamura H 1993 Structural studies of pressure-induced phase transitions in selenium up to 150 GPa. *Phys. Rev. B* **47**, 20–6

Batsanov S S 1997 Effect of high pressure on crystal electronegativities of elements. *J. Phys. Chem. Solids* **58**, 527–32

Brazhkin V V, Zerr A J 1992 Relative stability of red and black phosphorus at P < 1 GPa. *J. Mater. Sci.* **27**, 2677–81

Caldwell W A, Nguyen J H, Pfrommer B G, Mauri F, Louie S G, Jeanloz R 1997 Structure, bonding, and geochemistry of xenon at high pressures. *Science* **277**, 930–3

Car R, Parrinello M 1989 The unified approach for molecular dynamics and density functional theory. In: Polian A, Loubeyre P, Boccara N (eds.) *Simple Molecular Systems at Very High Density*, NATO ASI Series B: Physics. Plenum, New York, Vol. 186, pp. 455–76

Carlsson A E, Meschter P J 1995 Ab initio calculations. In: Westbrook J H, Fleischer R L (eds.) *Intermetallic Compounds. Principles and Practice*. Wiley, Chichester, UK, Vol. 1, pp. 55–76

Chen J H, Iwasaki H, Kikegawa T 1996 Crystal structure of the high pressure phases of bismuth BiIII and BiIII′ by high energy synchrotron X-ray diffraction. *High Press. Res.* **15**, 143–58

Chesnut G N, Vohra Y K 1998 Phase transformation in lutetium metal at 88 GPa. *Phys. Rev. B* **57**, 10221–3

Chesnut G N, Vohra Y K 1999 Structural and electronic transitions in ytterbium metal to 202 GPa. *Phys. Rev. Lett.* **82**, 1712–5

Daams J L C, Villars P, van Vucht J H N 1991 *Atlas of Crystal Structure Types for Intermetallic Phases*. ASM International, Materials Park, OH

Dabos-Seignon S, Dancausse J P, Gering E, Heathman S, Benedict U 1993 Pressure-induced phase transition in α-Pu. *J. Alloys Comp.* **190**, 237–42

Fujihisa H, Fujii Y, Takemura K, Shimomura O 1995 Structural aspects of dense solid halogens under high pressure studied by X-ray diffraction—molecular dissociation and metallization. *J. Phys. Chem. Solids* **56**, 1439–44

Hamaya N, Sakamoto Y, Fujihisa H, Fujii Y, Takemura K, Kikegawa T, Shimomura O 1993 Crystal structure of the distorted fcc high-pressure phase of praseodymium. *J. Phys. Cond. Matt.* **5**, L369–74

Hanfland M, Schwarz U, Syassen K, Takemura K 1999 Crystal structure of the high-pressure phase silicon VI. *Phys. Rev. Lett.* **82**, 1197–200

Holzapfel W B 1996 Physics of solids under strong compression. *Rep. Prog. Phys.* **59**, 29–90

Iwasaki H 1997 Comment on "bcc arsenic at 111 GPa: an x-ray structural study". *Phys. Rev. B* **55**, 14645–6

Iwasaki H, Kikegawa T 1997 Structural systematics of the high-pressure phases of phosphorus, arsenic, antimony and bismuth. *Acta Crystallogr. B* **53**, 353–7

Johansson B 1995 Crystal and electronic structure connections between the 4f and 5f transition metals. *J. Alloys Comp.* **223**, 211–5

Leigh G.J. (ed.) 1990 *Nomenclature of Inorganic Chemstry*, IUPAC Recommendations. Blackwell, Oxford, UK

Loubeyre P, LeToullec R, Hausermann D, Hanfland M, Hemley R J, Mao H K, Finger L W 1996 X-ray diffraction and equation of state of hydrogen at megabar pressures. *Nature* **383**, 702–4

Luo H, Greene R G, Ruoff A L 1993 β-Po phase of sulfur at 162 GPa: X-ray diffraction study to 212 GPa. *Phys. Rev. Lett.* **71**, 2943–6

Massalski T B 1990 *Binary Alloy Phase Diagrams*. ASM International, Materials Park, OH Vols. 1–3.

McMahon M I, Nelmes R J 1993 New high-pressure phase of Si. *Phys. Rev. B* **47**, 8337–40

Nelmes R J, Liu H, Belmonte S A, Loveday J S, McMahon M I, Allan D R, Häusermann D, Hanfland M 1996 *Imma* Phase of germanium at ~80 GPa. *Phys. Rev. B* **53**, R2907–R2909

Olijnyk H 1990 High pressure X-ray diffraction studies on solid N_2 up to 43.9 GPa. *J. Chem. Phys.* **93**, 8968–72

Parthasarathy G, Holzapfel W B 1988 High-pressure phase transitions in tellurium. *Phys. Rev. B* **37**, 8499–501

Porsch F, Holzapfel W B 1993 Novel reentrant high pressure phase transition in lanthanum. *Phys. Rev. Lett.* **70**, 4087–9

Reichlin R, McMahan A K, Ross M, Martin S, Hu J, Hemley R J, Mao H, Wu Y 1994 Optical, X-ray, and band-structure studies of iodine at pressures of several megabars. *Phys. Rev. B* **49**, 3725–33

Roof R B 1981 The crystal structure of high-pressure americium, phase III. *J. Appl. Crystallogr.* **14**, 447–50

Schulte O, Holzapfel W B 1993 Phase diagram for mercury up to 67 GPa and 500 K. *Phys. Rev. B* **48**, 14009–12

Schwarz U, Takemura K, Hanfland M, Syassen K 1998 Crystal structure of cesium-V. *Phys. Rev. Lett.* **81**, 2711–4

Söderlind P 1998 Theory of the crystal structures of cerium and the light actinides. *Adv. Phys.* **47**, 959–98

Takemura K, Fujihisa H 1993 High-pressure structural phase transition in indium. *Phys. Rev. B* **47**, 8465–70

Takemura K, Kobayashi K, Arai M 1998 High-pressure bct–fcc phase transition in Ga. *Phys. Rev. B* **58**, 2482–6

Takemura K, Shimomura O, Fujihisa H 1991 Cs(VI): a new high-pressure polymorph of cesium above 72 GPa. *Phys. Rev. Lett.* **66**, 2014–7

Tonkov E Y 1992 *High Pressure Phase Transformations: A Handbook*. Gordon and Breach, Philadelphia Vols. 1 and 2

Tonkov E Y 1996 *High Pressure Phase Transformations: A Handbook*. Gordon and Breach, Philadelphia Vol. 3

Villars P, Calvert L D 1991 *Pearson's Handbook of Crystallographic Data for Intermetallic Phases*. ASM, Materials Park, OH Vols. 1–4

Vohra Y K, Akella J 1991 5f Bonding in thorium metal at extreme compressions: phase transitions to 300 GPa. *Phys. Rev. Lett.* **67**, 3563–6

Vohra Y K, Akella J, Weir S, Smith G S 1991 A new ultra-high pressure phase in samarium. *Phys. Lett. A* **158**, 89–92

Vohra Y K, Beaver S L, Akella J, Ruddle C A, Weir S T 1999 Ultrapressure equation of state of cerium metal to 208 GPa. *J. Appl. Phys.* **85**, 2451–3

Winzenick M, Holzapfel W B 1996 Structural study of the high-pressure phase strontium III. *Phys. Rev. B* **53**, 2151–4

Winzenick M, Vijayakumar V, Holzapfel W B 1994 High-pressure X-ray diffraction on potassium and rubidium up to 50 GPa. *Phys. Rev. B* **50**, 12381–5

Young D A 1991 *Phase Diagrams of the Elements*. University of California Press, Berkeley, CA

Zeng W S, Heine V, Jepsen O 1997 The structure of barium in the hexagonal close-packed phase under high pressure. *J. Phys. Condens. Matter* **9**, 3489–502

Zhao Y C, Porsch F, Holzapfel W B 1994 Irregularities of ytterbium under high pressure. *Phys. Rev. B* **49**, 815–7

W. Steurer

Federal Institute of Technology, Zurich, Switzerland

Crystals, Dendritic Solidification of

Many crystalline materials grow in multibranched shapes, in which the branches form geometrical arrays that are directly related to the structure of the crystal. Salts crystallizing from aqueous solution tend to branch in this fashion and are often included in chemistry hobby sets. The branching often leads to a tree-like appearance, and the crystals are called dendrites, derived from the Greek *dendros* "tree." Spectacular geometric patterns can also be seen when ice crystals form on windowpanes in cold climates. It is less apparent that metals solidifying from the melt also form dendritic crystals. Nearly all of the cast metal products used in industry are composed of thousands to millions of tiny dendrites, which can be seen under the microscope on a specially polished surface. In metals the dendrites usually take a relatively simple form of evenly spaced arms branching off a single stem, as in Fig. 1(a). The size, shape, and orientation of the dendrites have major effects on the properties of a casting.

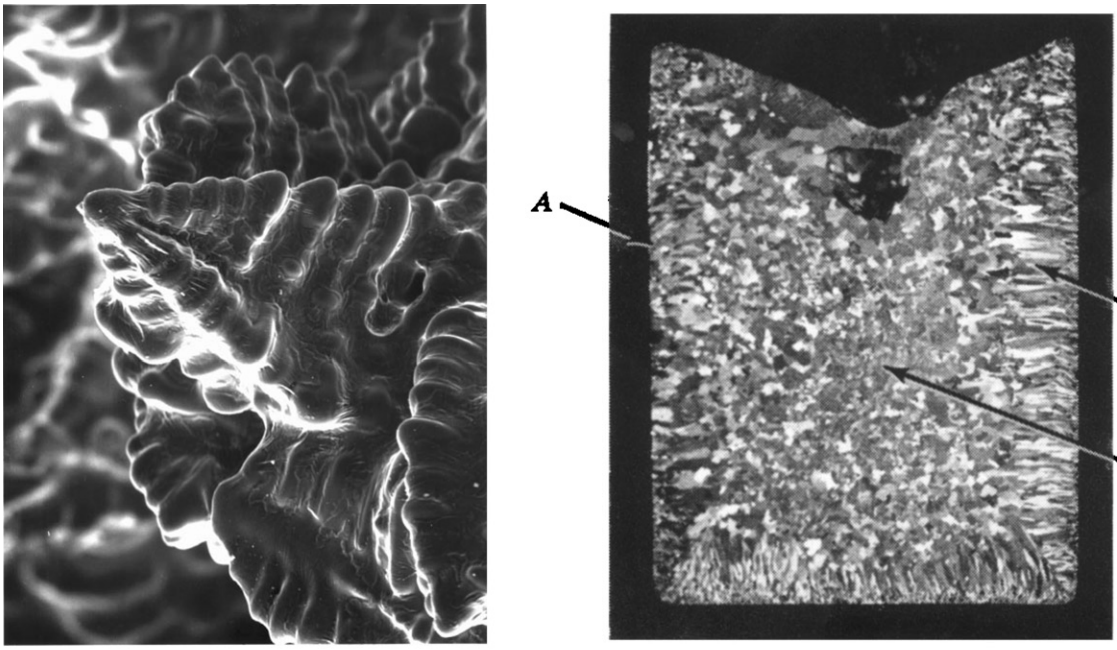
1. Dendrite Size and Orientation

Two types of dendrite, columnar and equiaxed, commonly form in a casting of a single-phase alloy, as in the simple ingot shape of Fig. 1(b). After pouring, a steep positive temperature gradient is established, the temperature increasing from the walls of the mold into the body of the melt. Crystals nucleate on the mold walls, forming equiaxed chill crystals (A). Competition suppresses unfavorably oriented grains and the survivors grow towards the center as parallel columnar grains (B). The gradient decreases progressively as heat is lost and may fall to zero in the central melt. Columnar growth then ceases and is replaced by central equiaxed grains growing dendritically (C). These are randomly oriented and, as each rejects latent heat, they form a negative temperature gradient in the surrounding melt.

1.1 Directional Solidification for Columnar Growth

Columnar growth can be simulated by directional solidification. Briefly, a long tube containing an alloy is moved through a furnace set at a temperature above the liquidus. At some distance along the specimen tube from the furnace maximum the temperature across the tube will be the liquidus temperature, making a solid–liquid interface. When the tube is moved towards the solid end of the alloy the interface will move into the liquid at a solidification rate determined by the rate of movement of the specimen tube. The positive temperature gradient in the melt at the interface can be set to a measured value so that changes in microstructure can be observed for varying values of initial composition, C_0, growth velocity, V, and temperature gradient, $G = dT/dx$, where x is measured as the distance from the solid–liquid interface.

Directional solidification can produce any of three types of microstructure which are distinguished in terms of the shape of the solid–liquid interface,



(a)

(b)

Figure 1
(a) Dendrites growing into a shrinkage cavity, and exposed as matrix liquid was withdrawn. The central part shows the progressive growth of side arms from a central stem. (b) Grain structure of a cast alloy ingot: A, fine chill crystals; B, columnar grains; C, central equiaxed zone. All grains are dendritic.

namely planar, cellular, and dendritic. The cell–dendrite transition is shown in Fig. 2. The transition between plane interface and cellular growth can be shown to follow the relationship:

$$\frac{G}{V} = \frac{mC_0}{D}\left(\frac{1-k}{k}\right) \tag{1}$$

where $G(^{\circ}\mathrm{C\,cm^{-1}}) = \mathrm{d}T/\mathrm{d}x$, $V(\mathrm{\mu m s^{-1}})$ is the growth rate, m is the slope of the liquidus line, D is the diffusion coefficient of solute in the melt, k is the distribution coefficient ($=C_S/C_L$) at temperature T (Fig. 3(a)), and C_0 is the original melt composition. Equation (1) is a straight line when G/V is plotted against C_0, as in Fig. 4. No such general relationship has yet been determined for the cell–dendrite transition, but a transition range as indicated in Fig. 4 can be observed experimentally.

2. *Constitutional Undercooling*

These changes can be understood in terms of constitutional undercooling theory, illustrated in Fig. 3, which depicts conditions in the melt adjacent to the solid–liquid interface for a single-phase alloy of composition C_0 (Fig. 3(a)). The interface will have depth, extending along the temperature gradient from liquidus to solidus temperature. In this range there will be a "mushy" zone, i.e., a mixture of liquid and solid. If the imposed temperature gradient is steep enough, equal to or greater than $\mathrm{d}T_A/\mathrm{d}x$ in Fig. 3(c), the mushy zone can be made so thin as to be effectively planar for a given ratio of G/V, as in Eqn. (1). For smaller G or larger V at given C_0 the conditions cross the line from planar to cellular in Fig. 4. The mushy zone expands and a steady-state condition can be reached with the solute distribution, as in Fig. 3(b). Solute rejected from the solid diffuses towards the end of the liquid column. Steady state is reached when $C_S = C_0$, i.e., when the solid forming at the interface has the same composition as the bulk melt and the melt composition at the interface is C_0/k. Figure 3(c) plots the liquidus temperature, T_L, for points on the curve in Fig. 3(b), reading from the phase diagram. In Fig. 3(c) the actual temperature, T_A, is shown as following a steep gradient, $G = dT_A/dx$, tangential to the T_L curve. If this condition is maintained, the interface at $x=0$ must be planar, because any protrusion of solid into the liquid will be at a temperature above T_L and will melt off. However, in Fig. 3(d) the applied temperature gradient is lower and the melt temperature, T_A, is below T_L over the distance X_1 from the interface. Thus, there is a film of liquid of

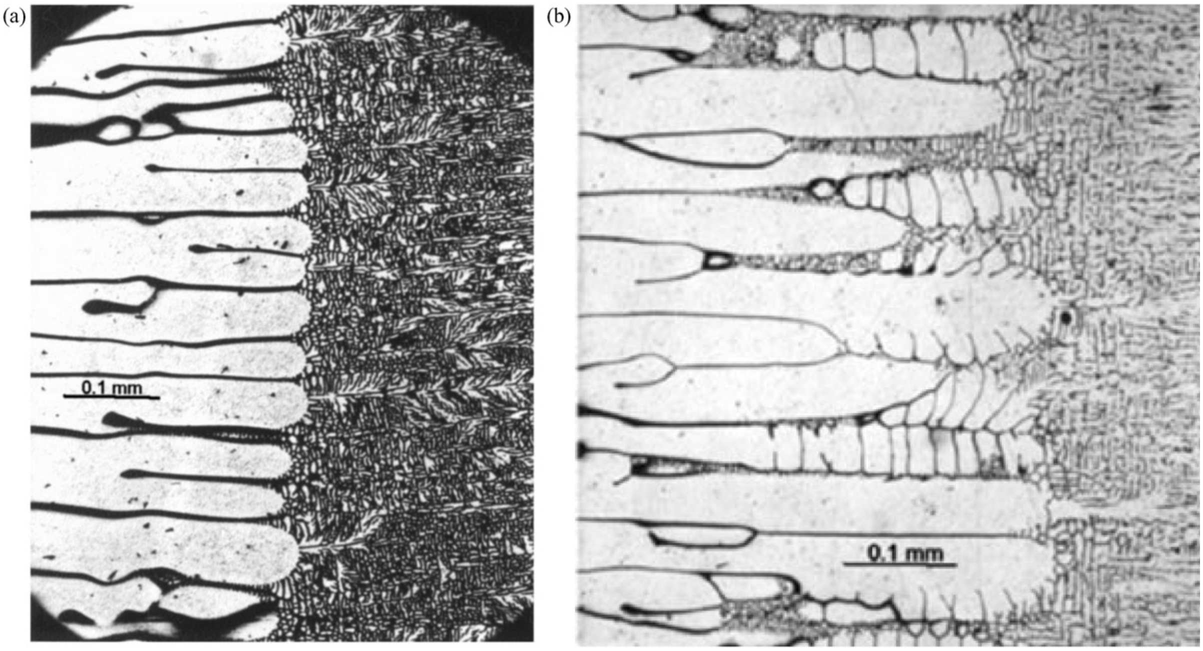


Figure 2
Aluminum–copper alloys quenched during directional solidification to reveal solid–liquid interfaces. Growth direction is left to right. (a) Cellular interface grown at low velocity; (b) breakdown to cellular dendritic at higher velocity. In each case the fine-branched structures on the right-hand side are dendrites grown at high velocity from liquid thermally undercooled by the quench (after Sharp and Hellawell 1970).

thickness X_1 adjacent to the interface which is undercooled relative to the liquidus temperature. This is called *constitutional undercooling* to indicate that the undercooling arises from a change in composition, not temperature. For practical purposes the thickness, δ_c, of the boundary film is given by the following (Kurz and Fisher 1989):

$$\delta_c = \frac{2D}{V} \tag{2}$$

Under this condition, any protuberance forming on the interface at $x = 0$ would find itself in undercooled liquid and would survive. An unstable interface with many perturbations (Kurz and Fisher 1989) quickly settles into a regular pattern called cells (Fig. 2(a)). If undercooling increases the cell tips grow further into the liquid and develop side branches to give the tree-like morphology of dendrites (Fig. 2(b)). The transition from cells to dendrites is not sharp and a transition range is indicated in Fig. 4.

The significance of cell formation is that it provides a means of packing away the solute rejected from the solid, so long as there is not too much of it. Instead of the solute diffusing forward into the bulk liquid, it diffuses across the interface to the cell boundaries and is deposited there. In transverse section the cells usually take up a hexagonal pattern, and the boundaries of the hexagons are occupied by high-solute material which can be two-phase on final solidification.

3. *Dendrite Formation in Alloys*

3.1 Columnar Grains

In alloys of higher solute content or at higher growth velocities the rate of rejection of solute increases and the system must adopt more drastic means to dispose of it. The cells are replaced by a smaller number of long spines of low solute content that grow rapidly into the melt (Fig. 2(b)), leaving relatively large channels of solute-rich melt between them. As the spines thicken and solute is rejected laterally the conditions along the flanks of the spines are much the same as those at the tips, with low-solute solid in contact with high-solute liquid (Fig. 3(d)), so much the same mechanism operates. Regularly spaced branches grow into the solute-rich liquid to form side arms in directions determined by the crystal structure of the solid. In cubic crystals the side arms are at right angles to the central spines.

The rejection of high-solute liquid to the spaces between dendrite arms is known as *microsegregation*.

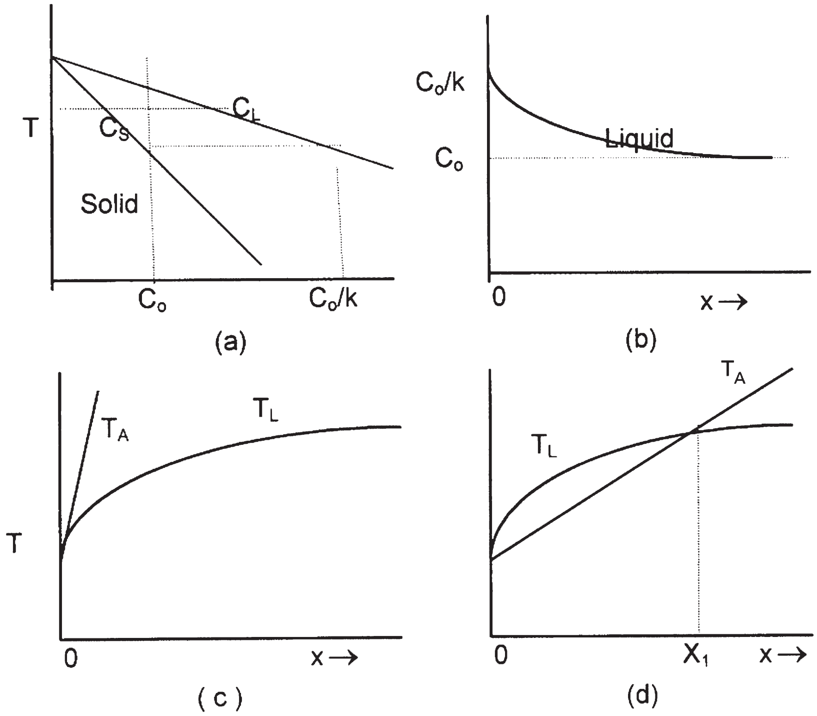


Figure 3
Conditions at the solid–liquid interface during directional growth of a dilute metallic alloy. (a) Portion of a phase diagram showing bulk alloy composition, C_0, and compositions C_L and C_S at an isotherm between liquidus and solidus. (b) Diffusion curve at interface during steady-state growth with a planar interface. (c) Temperatures T_A (applied) and T_L (liquidus) for planar growth of an alloy. (d) Constitutional undercooling condition leading to cellular or dendritic growth.

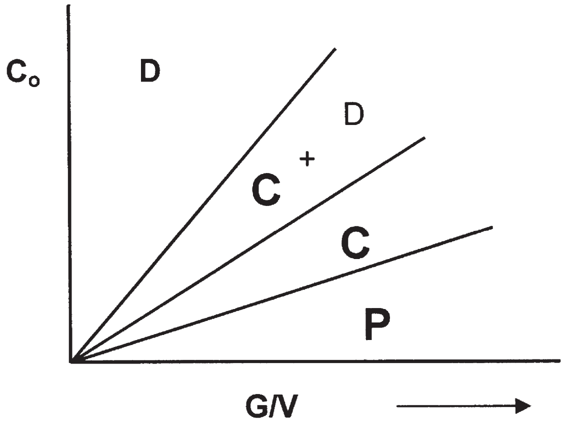


Figure 4
Changes of solid–liquid interface shape from planar (P) to cellular (C) to dendritic (D) as alloy content, C_0, increases or G/V decreases.

In the last stages of freezing some of this enriched liquid is squeezed towards the walls of the mold or towards the central melt. This is *macrosegregation*.

3.2 Equiaxed Dendrites

As an ingot freezes and heat is lost to the mold, the gradient dT_A/dx at the interface decreases until it may become nearly zero across the remaining melt in the center of the ingot. While columnar growth continues the solubility curve T_L will retain its shape (Fig. 3(d)) while the T_A curve is almost horizontal, with the result that strong constitutional undercooling extends through the whole of the remaining melt. Nucleation can then occur readily on any nucleant species that may be present and subsequent growth into the undercooled melt will be dendritic. In marked contrast to constrained directional freezing, growth is multidirectional, so that primary stems of cubic crystals (for example) can grow in all six $\langle 100 \rangle$

directions from the origin. The space between will be filled by secondary arms. The temperature gradient in the melt surrounding each grain is negative, being sustained by latent heat emissions from the solid–liquid interface, and growth velocity will be as fast as the undercooling permits. Heat loss is through the melt until the crystals impinge upon each other and on the columnar growth front, giving continuous paths through to the solid. Equiaxed growth can be induced throughout a casting if a highly efficient nucleant species is available, as is the case for aluminum alloys.

4. Dendrite Formation in Pure Metals

There can be no constitutional undercooling in a pure melt, since there is no solute rejection. Dendrites can form only in a thermally undercooled melt, which can occur when the melt is poured into a cold mold, in which case growth occurs into a negative gradient, as for equiaxed alloy dendrites. However, growth in a casting will be wholly columnar, since equiaxed dendritic growth is not possible, and when thermal undercooling is dissipated the columnar front must become planar. In practice, nominally pure metals usually contain some impurity content and the columnar growth is likely to be cellular.

5. Dendrite Arm Spacing

A convenient measure of the effects of solidification conditions on dendrite structure is dendrite arm spacing, which is the average spacing between dendrite arms. Secondary dendrite arm spacing is usually the most useful parameter. Numerous experiments have confirmed that secondary spacing, d_s, is closely dependent on the total time, t_f, during which liquid and solid have coexisted:

$$d_s = at_f^n \tag{3}$$

where a and n are constants for specific conditions. Measured values of n range between 0.25 and 0.4, the lower values relating to earlier stages of dendrite growth, when the fraction of solid, f_s, is small, while higher values occur when f_s is high (Dann *et al.* 1979). As an alloy cools between liquidus and solidus the secondary arms tend to absorb each other as they grow into the space between primary stems, so that the number of secondary arms per unit distance along a primary stem decreases as time of contact between liquid and solid increases. Thicker arms tend to grow at the expense of thinner neighbors by one of three recognized mechanisms. All three operate by diffusion from surfaces of higher curvature to those of lower curvature, as in the Ostwald ripening of precipitates. A further mechanism occurring at higher fractions of solid, as the dendrite arms thicken, is coalescence of adjacent arms, and this is thought to account for the higher rate of coarsening at high f_s (Dann *et al.* 1979).

Local solidification time, t_f, is related to cooling rate:

$$t_f = \frac{\Delta T_S}{GV} \tag{4}$$

where ΔT_S is the temperature range between liquidus and final solidification, G is the temperature gradient, V is the growth velocity, and GV has units of cooling rate. It follows from Eqns. (3) and (4) that dendrite arm spacing can be used as a measure of cooling rate and this has proved valuable where cooling curves are not available. The method is important in rapid solidification experiments when direct measurement of cooling rate may be difficult or impossible, and in studies of castings in which cooling rate will vary with position in the casting.

6. Modeling of Dendritic Growth

Attempts to develop models to predict dendrite form and behavior under given conditions have given rise to a substantial body of literature. Mathematical models involve simultaneous solutions of the diffusion equation for heat and solute rejected into the liquid from the curved tip of a dendrite, while the curvature is itself affected by the heat and solute distribution. This interaction makes exact theories difficult, although close approximations have been developed for directionally solidified dendrites (Trivedi 1980, Hunt 1984).

Most analyses start with a needle dendrite, which is a single dendrite stem growing into undercooled melt, with a tip shape compatible with diffusion solutions (Huang and Glicksman 1981). The driving force for growth of the dendrite can be expressed as undercooling of the melt in contact with the tip below the liquidus temperature for the bulk alloy. The total undercooling, ΔT, can be divided into three components:

$$\Delta T = \Delta T_r + \Delta T_c + \Delta T_t \tag{5}$$

where ΔT_r is the undercooling due to curvature. The temperature of equilibrium between solid and liquid is lowered below that for a planar interface in inverse proportion to the radius of curvature, r, of the tip:

$$\Delta T_r = K\Gamma \tag{6}$$

where K, the curvature, is $2/r$ for a hemispherical surface and Γ is the Gibbs–Thomson coefficient, which can be related to surface energy, σ, by:

$$\Gamma = \frac{\sigma}{\Delta S_f} \tag{7}$$

where ΔS_f is the entropy of melting of the solute phase.ΔT_c is the solute undercooling owing to

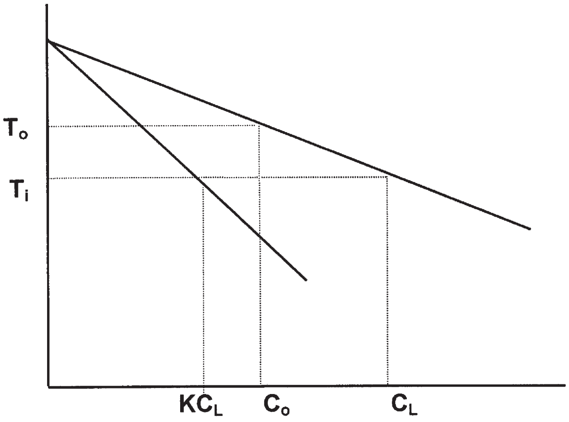


Figure 5
Solute undercooling related to melt composition of the dendrite tip during growth of a needle dendrite.

supersaturation, Ω, of the melt at the interface with respect to the solute. Referring to Fig. 5, $\Omega = (C_L - C_0)/(C_L - kC_L)$ and

$$\Delta T_c = T_0 - T_i \tag{8}$$

ΔT_t is the thermal undercooling, which may exist in the bulk melt ahead of the dendrite tip owing to heat extraction prior to the start of solidification.

Analytical models attempt to solve Eqn. (5) in terms of measurable quantities. Thus supersaturation can be expressed as $Vr/2D = r/\delta_c$ (Eqn. (2)). This dimensionless group is frequently written as P_c, the Peclet number for solute diffusion. If P_c is constant, V and r are interdependent and an additional expression is needed to predict tip radius at a given velocity. Detailed analysis tends to be concentrated on the nature of this expression (Laxman 1985).

The factors discussed above apply to a free needle dendrite growing into thermally undercooled melt. In directional solidification a growing dendritic crystal appears as a parallel array (Fig. 2(b)) and the factors determining the spacing between dendrite tips remain controversial (Hunt and Lu 1996). Further, the sides of the dendrite tip suffer a high frequency of perturbations which coarsen into dendrite arms (Huang and Glicksman 1981), as discussed in Sect. 6. Finally, analysis of the interaction of heat and solute diffusion during growth of an equiaxed dendrite is more complex again, and exact prediction remains a challenge. Numerical modeling is used to make realistic predictions (Hunt and Lu 1996, Nastac 1999).

7. *Summary*

Dendrites in metals are multibranched crystals formed during solidification. In alloy castings distinction is made between columnar and equiaxed dendrites, which differ in growing into positive and negative temperature gradients, respectively. Columnar dendrites grow at a rate determined by the constitutional undercooling of the melt at their tips. Equiaxed dendrites grow as fast as the total melt undercooling permits. Pure metals cannot form equiaxed crystals, since constitutional undercooling cannot occur in the absence of alloying elements. Secondary dendrite arm spacing increases with time of contact between liquid and solid. Analytical modeling of dendrite growth is not completely successful, but close prediction of dendrite shape and spacing can be achieved by numerical modeling.

Bibliography

Chalmers B 1964 *Principles of Solidification*. Wiley, New York
Dann P C, Hogan L M, Eady J A 1979 Mechanisms of coarsening of secondary dendrite arm spacings in Al–Cu alloys and an organic analogue. *Met. Forum* **2**, 212–9
Flemings M C 1974 *Solidification Processing*. McGraw-Hill, New York
Huang S C, Glicksman M E 1981 Fundamentals of dendritic solidification: I. Steady-state growth. *Acta Metall.* **29**, 701–15
Hunt J D 1984 Steady state columnar and equiaxed growth of dendrites and eutectic. *Mater. Sci. Eng.* **65**, 75–83
Hunt J D, Lu S-Z 1996 Numerical modelling of cellular-dendritic array growth: spacing and structure predictions. *Metall. Mater. Trans.* **27A**, 611–23
Kurz W, Fisher D J 1989 *Fundamentals of Solidification*, 3rd edn. Trans Tech, Aedermannsdorf, Switzerland
Laxmanan V 1985 Dendritic solidification: III. Some further refinements to the model for dendritic growth under an imposed temperature gradient. *Acta Metall.* **33**, 1475–80
Nastac L 1999 Numerical modelling of solidification morphologies and segregation patterns in cast dendritic alloys. *Acta Mater.* **47**, 4253–6
Sharp R M, Hellawell A 1970 Solute distributions at non-planar solid–liquid growth fronts. *J. Cryst. Growth* **6**, 253–60
Trivedi R 1980 Theory of dendritic growth during the directional solidification of binary alloys. *J. Cryst. Growth* **49**, 219–32

L. M. Hogan
University of Queensland, Brisbane, Queensland, Australia

Cuticle

The outer covering of any arthropod (“jointed-limbed” animal) is called the *cuticle*. It is primarily made of chitin fibers in a more or less hydrated protein matrix. It performs a dual service as a barrier layer, separating outside from inside, and as a skeleton. It therefore needs the properties of a selective membrane and has to be stiff and strong enough to carry loads, both exogenous and endogenous. It is a single layer of material, highly convoluted and sculpted, produced by a single layer of cells—the epidermis. It can vary in thickness from 1.0 µm, covering the external gills of aquatic larvae, to 200 µm, in the unstretched extensible cuticle in the abdomen of the female locust. Since the cuticle is mostly stiff and is never entirely extensible, the arthropod has to slough off its old cuticle (ecdysis) having grown a new, folded, cuticle beneath, which it can inflate and harden and thus allow for growth. In the account that follows most of the information derives from work on insects. A general source is Binnington and Retnakaran (1991).

1. *Morphology of Cuticle*

The outermost layer is the epicuticle (Fig. 1), about 0.5 µm thick, most probably made of phenolically tanned protein. The epicuticle is very stiff and inextensible and forms the template for the next stage (or instar) in an insect's development, in that it limits the expansion of the insect after ecdysis. It is produced as separate small plaques on the surface of the epidermal cells, and immediately poses a problem. Since this layer defines the outward appearance of the insect, how are shape and size controlled at the level of the individual epidermal cells?

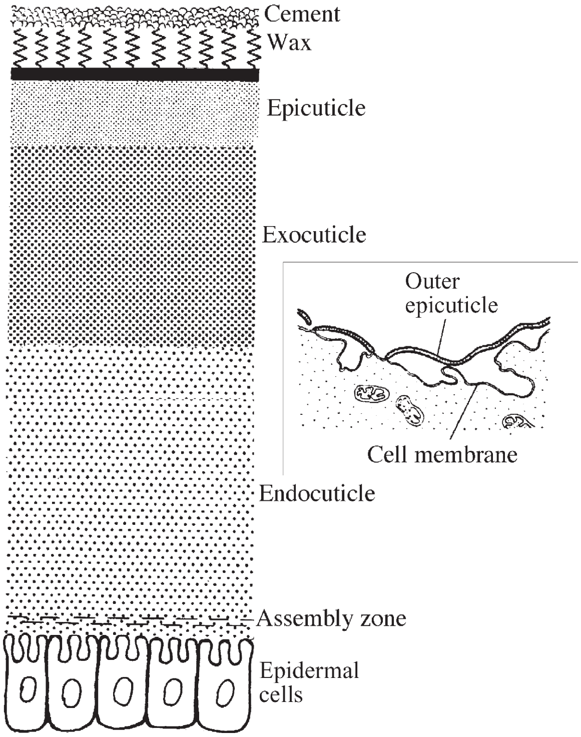


Figure 1
Major layers in typical insect cuticle. The epicuticle, cement, and wax layers are much thinner in proportion. The first layer to be laid down by the epidermal cells is the outer epicuticle (detail, right), starting as isolated sheets which combine to define the size and shape of the insect (modified by permission of The Royal Entomological Society from Weis-Fogh 1970).

The cuticulin is normally covered by a layer of wax, which is the main waterproofing barrier, and this layer in turn can be protected by a film of (probably) polymerized lipid which is called the cement layer. The cement and wax layers tend to be removed by environmental abrasion, especially in soil-living insects, but can be renewed from the epidermal cells which secrete material in solution through a series of holes (pore canals) extending through the thickness of the cuticle.

Beneath the epicuticle is the main structural part of the cuticle, which can be divided into two further layers. The outer (exocuticle), is chemically very inert, being relatively dehydrated and cross-linked. It is therefore also stiff, tends to brittleness, and is insoluble in water. The inner layer (endocuticle) is more highly hydrated and softer and more readily soluble, especially in dilute buffer or hydrogen-bond breakers. The thickness of these layers may be considered to vary during the growth and development of the cuticle, since endocuticle is secreted and resorbed by the epidermal cells, and if the outer part of the endocuticle (abutting the exocuticle) is tanned (see Sect. 2.3) it becomes included with the exocuticle. Hard or stiff cuticle is largely exocuticle (i.e., highly tanned) and forms the main plates (sclerites) and tubes of the skeleton. Soft cuticle provides the flexible jointing (arthrodial membrane) between the stiff parts and is composed almost entirely of (hydrated) endocuticle. It is also called intersegmental or pleural membrane depending on where it occurs.

Running through the thickness of the cuticle there is commonly an array of fibers (tonofibrils) which are more apparent at muscle insertions and cuticle which can be highly deformed (Neville 1975).

2. *Chemistry of the Components*

2.1 *Chitin*

Chitin is a fairly completely acetylated polysaccharide akin to cellulose. It is produced from a polymerase which is embedded in stainable patches (plaques) in the cell membrane at the tips of microvilli ("little fingers") extending from the surface of the epidermal cells (Fig. 2). The zone of cuticle immediately above the epidermal cells, into which the chitin is secreted, is variously known as the Schmidt layer or the assembly zone. It is highly hydrated, containing about 90% water (Weis-Fogh 1970), presumably allowing the chitin chains to adjust their positions, and present the liquid crystalline structures which are so typical of cuticles (Neville 1993). The sugar residues

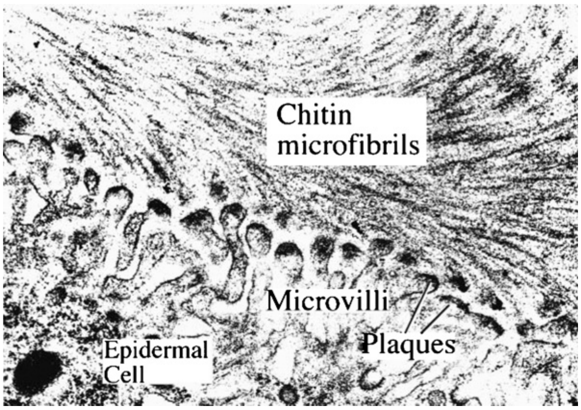


Figure 2
Microvilli producing chitin from plaques on their tips (× 45 500).

in the chitin molecule are heavily hydrogen bonded, making it very stiff and stable. The chains are arranged antiparallel (Neville 1975). Within the body of the cuticle they are assembled into microfibrils about 3nm in diameter, each containing 18–20 molecular chains and about 0.3 μm long. The stiffness of chitin cannot be less than that of cellulose, for which the best estimate is currently about 130 GPa, still only half the theoretical stiffness of 250 GPa.

2.2 *Proteins*

Protein is synthesized within and secreted from the epidermis, interacting with the chitin in the Schmidt layer. The protein has to produce a matrix of varying mechanical properties, which will also interact with and stabilize the chitin. The stiffness of the cuticle is then due to the extent of interaction of the protein with the chitin, with other proteins and internally with itself. Stiffness can vary from a kilopascal in the extensible intersegmental membrane of the locust to a few gigapascals in well-tanned cuticles (Vincent 1980). The interaction of the proteins with the chitin seems to be fairly consistent in that even in the softest of cuticles (in which the interactions are least developed) a strong solvent (e.g., 5% NaOH at 100 °C) is still required to remove the protein from the chitin. For the rest of the protein fraction, in most cuticles the interactions are mediated by water content, the water acting as a plasticizer. The proteins vary widely in hydrophobicity (Andersen *et al.* 1995). In general, proteins from more highly evolved insects tend to be less hydrophobic, even though they will form stiff cuticles in the same way. Perhaps their tendency to form secondary structures is more highly developed. Certainly such proteins very readily form beta structures when dehydrated.

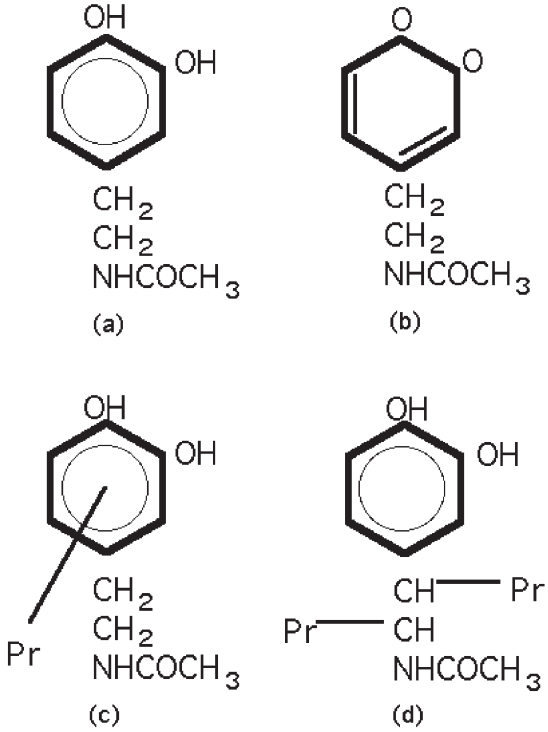


Figure 3
Highly summarized schema showing (a) *N*-acetyldopamine (one of the main tanning precursors) which is oxidized to the quinone form (b). This can then cross-link with proteins either through the aromatic nucleus (c) or the side-chain (d).

2.3 *Control of Stiffness*

There is much confusion over the main cause of cuticular stiffening, a process variously called *sclerotization* or *tanning* or, colloquially, "hardening and darkening." The only indisputable event is the introduction of *o*-dihydroxyphenolic compounds (Fig. 3) into the cuticle when the insect hardens or stiffens its cuticle after ecdysis. They are converted into the more reactive quinone form within the cuticle. They very probably interact covalently with the protein, but the degree to which they cross-link it is not known, and there is no direct evidence that they stiffen the cuticle (Andersen *et al.* 1995). However, this stiffened material can often be softened in boiling water or even made rubbery by soaking it in formic acid. This indicates that the secondary reactions made possible by the removal of water from the protein are much more important than the phenols in stiffening the cuticle, and that the control of stiffness is, in general, a matter of manipulating the water content. In some cuticles with reversible stiffness, the insect can increase the water content so that the modulus decreases. This happens in the blood-sucking bug *Rhodnius*, for instance, which can change the pH of the cuticle from about 7 to below 6, thereby increasing cuticular water content from about 26% to 31%, dropping its stiffness from 250 MPa to 10 MPa and increasing its extensibility from 10% to more than 100%.

It has to be said that the concept of water control as the critical function in the stiffness and tanning of cuticle does not command a wide following at present. Although it was noted as an important factor early on, it has been largely ignored except by a few people (Vincent and Hillerton 1979, Andersen *et al.* 1995, 1996). There is some good evidence that the phenolic residues are covalently bonded to the proteins (Andersen *et al.* 1996). What is lacking is an understanding of how much this cross-linking contributes to the stiffness and stability of the cuticle.

3. Other Components

There is some evidence for lipid through the thickness of the cuticle though its mechanical functions are uncertain. Some flies (Diptera) have calcium in the puparium (the tanned larval skin surrounding the delicate pupa within which the adult insect develops from the maggot), which presumably stiffens and strengthens this protective structure. Insects as a whole do not incorporate calcium into their cuticle since the weight penalty is too great for an arthropod that flies. However, it is the main stiffening agent in the cuticle of crustacean arthropods (crabs, shrimps, krill, etc.) where weight is not an issue.

4. Cuticle as a Fibrous Composite

In general, cuticle can be modeled as a fibrous composite, although few have done the calculations (Vincent 1980). In simple terms it can be shown that the stiffness of the cuticle is greater parallel to the orientation of the chitin microfibrils (Barth 1973). Conversely, the stiffness can be greatly reduced orthogonal to the chitin orientation.

5. Resilin

Resilin was identified as a specialized form of cuticle by Weis-Fogh, and subsequently shown to have excellent rubbery properties which aid energy conservation in flight (Neville 1975). Modern studies suggest that at least parts of the protein have a defined (nonrandom) conformation, so not all of the molecule can be random, as Weis-Fogh deduced. This topic therefore needs to be revisited, since the mechanical properties predicted quite well the degree of cross-linking, but only on the basis of rubber elasticity.

6. Surface Hardening

Some areas of the cuticle, notably the cutting surfaces of the mandibles, and also the egg-laying tube (ovipositor) of certain wasps contain up to 10% by weight of zinc or manganese. Their chemistry and mode of incorporation are unknown, but they can double the surface hardness of the cuticle, thus making it more resistant to abrasion and wear. These metals have also been found in the mandibles of spiders and some crustacea. There is much evidence to suggest that the metal has chloride as the counter-ion, and that they are incorporated into phenolically tanned proteins throughout the groups of "lower" animals.

7. Localized Structures

The cuticle is the skeleton, and so has to provide joints and rubbing surfaces. These have not been studied quantitatively. It also has to transmit information about the inside world to the sensitive organism living within. It has to do this without becoming weakened, either by the presence of holes or by the existence of sudden changes in mechanical properties. Mechanosensory hairs sit in resilin-lined holes (sockets) which must not weaken the skeleton. There are strain sensors (campaniform sensilla in insects, lyriform slit sensilla in spiders) which are basically holes which are allowed to deform—the sensing involves measuring the amount by which the holes deform when the surrounding skeleton is loaded (Barth and Pickelmann 1975). The holes are either too small to pose a problem or the chitin microfibrils are deformed around the hole, taking the forces with them and so effectively making the hole less visible to the rest of the structure.

Paradoxically this also increases the sensitivity of the holes to external strains, effectively separating stress concentration and strain amplification. Muscle insertions pose particular problems, representing as they do a classical problem in the orthogonal joining of fibrous composites. The trick is fibrous continuity from the muscle fibers through the epidermal cells, by means of thickened membranes on both inside (basal) and outside (apical) membranes of the cell, with

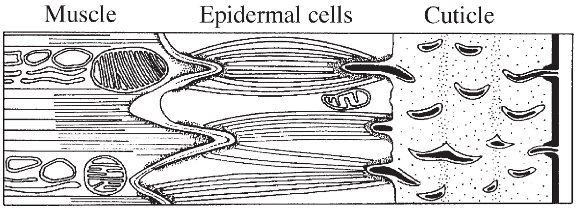


Figure 4
Mechanical continuity at a muscle attachment. The forces are fed from the muscles via microtubules in the epidermal cells to tonofibrillae which spiral through the cuticular layer (adapted by permission of The Company of Biologists Ltd. from Caveney 1969).

microfilaments feeding the forces smoothly into microtubules ($G = 2\,\mathrm{GPa}$) which take the force across the cell. The forces are fed into the cuticle via tonofibrils which emerge from pits in the apical membrane and traverse the cuticle orthogonal to the chitin microfibrils in the bulk of the cuticle (Fig. 4, Caveney 1969).

8. *Color*

Cuticle can be colored by pigments or by diffraction and interference effects allied to controlled reflectivity. In some scarab beetles, more or less total reflection of incident light is achieved by manipulating the polarization of the light as it passes through, and is reflected by, layers in the cuticle. Black pigmentation may have mechanical implications, since it is achieved by melanin particles which could also reinforce the cuticle by a filler effect. Significantly the knees, mandibles, and claws of insects, which are likely to experience greater forces than the rest of the cuticle, are often black.

Interference layers are present in lepidoptera, especially the cuticle of the quiescent or pupa stage which is also known as the chrysalis from the Greek *chrysos*, meaning golden. They are also present in adult beetles such as the metallic green tiger beetle. Many insects have diffraction gratings on the surface of the cuticle, especially day-flying lepidoptera. The varying color, as the wing of the flying insect changes its position relative to sunlight falling on it, makes it difficult for a bird (the main predator of insects) to follow.

9. *Pesticides*

Since chitin is almost diagnostic of arthropods (otherwise it occurs only in fungi) it makes an easy target for pesticides. In the late 1960s a Dutch firm, Duphar, stumbled upon a chemical (later named Dimilin) effectively made from two herbicide molecules which totally inhibited the polymerization of chitin. It has been followed by other pesticides with similar action (Binnington and Retnakaran 1991). Fed to developing insects it stops the proper development of the cuticle, so that the newly emerging insect bursts as it swallows air to expand itself to its new size for the next stage in its life.

10. *Permeability*

Insect cuticle has some intriguing permeability properties. Many small insects, terrestrial as well as aquatic, respire to a significant degree through the general body surface as well as using the breathing tubes (tracheae) which so typify arthropods. Some desert-living insects can absorb water from humid air using a combination of capillary condensation between very fine hairs and changing the hydrophilicity of the cuticle by controlling its charge density with a salt solution (O'Donnell 1982). This is an extreme example of a general phenomenon. The wax layer of the cuticle not only protects the insect from desiccation but forms a barrier for penetration of topically applied pesticides, which have to cross this hydrophobic boundary and then enter the hydrophilic environment of the endocuticle and the body cavity of the insect.

Permeability is also important for chemosensing. The hairs on antennae and other sense organs have minute pores in the end so that the chemical to be sensed can contact cell membrane more or less directly. There is no general dermal chemical sensitivity since the exocuticle is chemically so inert.

11. *Biomimetics*

Such a versatile and multifunctional material as insect cuticle can hardly avoid the predatory eyes of the copyist. Strictly speaking we would wish to know what the design principles of insect cuticle might be. As far as can be seen, it follows known concepts of composite theory, except that it is more complex and its properties are more finely and more locally controlled than anything which can be synthesized by us at present. Its chemistry is more complex and more varied, but until its mechanical properties have been measured with greater precision and analyzed using composite theory, it is impossible to say whether there is anything there which is worth copying. Even so, the integration of function, the multilayer construction, and the design of integral sensing are well worth study since we have no technology like this, and seeing how it is done by the masters of fibrous composite materials will undoubtedly enable us to design more versatile structures. Making them will be another matter!

Bibliography

Andersen S O, Hojrup P, Roepstorff P 1995 Insect cuticular proteins. *Insect Biochem. Mol. Biol.* **25**, 153–76

Andersen S O, Peter M G, Roepstorff P 1996 Cuticular sclerotization in insects. *Comp. Biochem. Physiol. B* **113**, 689–705

Barth F G 1973 Microfiber reinforcement of an arthropod cuticle laminated composite material in biology. *Z. Zellforsch* **144**, 409–33

Barth F G, Pickelmann P 1975 Lyriform slit sense organs. Modelling an arthropod mechanoreceptor. *J. Comp. Physiol.* **103**, 39–54

Binnington K, Retnakaran A (eds.) 1991 *Physiology of the Insect Epidermis*. CSIRO, Melbourne

Caveney S 1969 Muscle attachment related to cuticle architecture in apterygota. *J. Cell Sci.* **4**, 541–59

Neville A C 1975 *Biology of Arthropod Cuticle*. Springer, Berlin

Neville A C 1993 *Biology of Fibrous Composites; Development Beyond the Cell Membrane*. Cambridge University Press, Cambridge, UK

O'Donnell M J 1982 Hydrophilic cuticle—the basis for water vapour absorption by the desert burrowing cockroach, *Arenivaga investigata*. *J. Exp. Biol.* **99**, 43–60

Vincent J F V 1980 Insect cuticle—a paradigm for natural composites. *Symp. Soc. Exp. Biol.* **34**, 183–210

Vincent J F V, Hillerton J E 1979 The tanning of insect cuticle— a critical review and a revised mechanism. *J. Insect Physiol.* **25**, 653–8

Weis-Fogh T 1970 Structure and formation of insect cuticle. In: Neville A C (ed.) *Insect Ultrastructure*. Royal Entomological Society, London, pp. 165–85

J. F. V. Vincent
University of Reading, UK

Cyclic Polymers, Crystallinity of

The crystallinity of large cyclic polymers is similar to that of their linear analogs, the effect of cyclization being unimportant in the highly-folded semicrystalline state. Accordingly, the general effects in the crystallization and crystallinity of large cyclics are effectively covered elsewhere. However, if the chains are short, and particularly if the chains are uniform in length (oligomers), then significant differences in structures and properties emerge, study of which can give valuable information on chain folding. Figure 1 represents uniform cyclic oligomers packed into a layer crystal in extended conformation with parallel-packed stems and tight folds; the stem length necessarily being half that of the skeletal length. The corresponding model for cyclic polymers, having a distribution of sizes crystallized into a lamella stack, would include a disordered lamella-surface layer and, at large size, folding of the rings.

If the chain length in a cyclic oligomer is very short then different structures will be formed, giving x-ray diffraction (XRD) patterns clearly distinct from those of high polymers. The ring size at the transition to the polymer-like structure is an important consideration. The evidence available points to this transition occurring in the range 50–150 skeletal atoms.

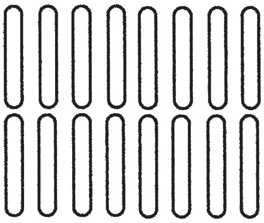


Figure 1
Schematic diagram showing uniform cyclic oligomers packed into a layer crystal in extended conformation with parallel-packed stems and tight folds.

Progress in understanding the physicochemical properties of cyclic polymers has been founded upon advances in preparative chemistry and in high-resolution separation methods capable of yielding uniform cyclic oligomers and cyclic polymers with narrow chain-length distributions. This aspect is well-described in Semlyen's two compilations (Semlyen 1985, 2000). Of the methods available, ring-closure of an α,ω-difunctional chain at high dilution (dilution method) and extraction of cyclics from a ring-chain equilibrate are both well established. A recent development, useful for many commercially-important polymers (e.g., polyesters, polycarbonates, polysulfones) is the generation of cyclics by cyclodepolymerization (e.g., Baxter *et al.* 1998). The studies summarized in Sects. 1 and 2 are chosen to illustrate some of the points made above.

1. *Cyclic Oligomers*

1.1 *Cyclic Alkanes*

Large cycloalkanes were prepared from lengthy α,ω-alkanediynes by the dilution method (Lee and Wegner 1985). The subsequent studies of the crystallinity of a wide range of c-alkanes (c-C_{24}–c-C_{288}) involved a wide range of techniques e.g., wide- and small-angle x-ray scattering (WAXS and SAXS), electron diffraction, Raman, and IR spectroscopy, differential scanning calorimetry (DSC) (Lieser *et al.* 1988). A variety of crystal forms (monoclinic, orthorhombic, hexagonal), all with parallel stems, were confirmed. When crystallized from the melt, the orthorhombic form (with the "polyethylene" sub-cell) was unstable for c-C_{72} and c-C_{96} but stable for c-C_{144} and c-C_{288}, pointing to the latter as good models for adjacent chain folds in the linear polymer.

1.2 *Cyclic Oligo(oxyethylene)s*

Large cyclic oligo(oxyethylene)s (range c-E_{15} to c-E_{27} where E represents an oxyethylene unit, OCH_2CH_2, 45–81 skeletal atoms) were prepared from α,ω-hydroxyoligomers by the dilution method (Yang *et al.* 1997). Results from WAXS, SAXS, and Raman spectroscopy were combined to show that the crystal structure typical of high-molar-mass poly(oxyethylene) (POE), i.e., with a monoclinic sub-cell containing four parallel helical chains of alternating handedness, was stable for melt-crystallized c-E_{18} and c-E_{27}, i.e., the transition to the "POE" structure was at approximately 50 skeletal atoms. Thermal analysis (DSC) showed that the enthalpy of formation of the fold surface in the layer crystals of large cyclics was 36 kJ (mol of folds)$^{-1}$ compared with just 4 kJmol^{-1} for methyl-terminated chain ends (8 kJmol^{-1} for two ends) in the layer crystals of corresponding linear

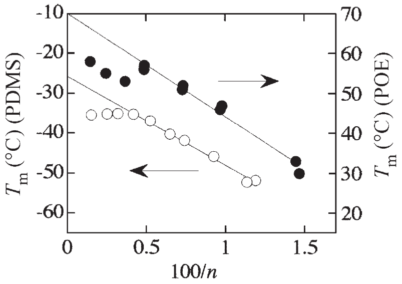


Figure 2
Dependence of melting temperature on inverse ring size n for (○) PDMS and (●) POE fractions after rapid crystallization. n is the number of skeletal atoms.

oligomers. A similar surface enthalpy (ca. 36 kJmol^{-1}) should hold for adjacent-folds in crystalline POE.

2. Cyclic Polymers

2.1 Cyclic Poly(dimethylsiloxane)

Narrow fractions of cyclic PDMS (84–682 skeletal atoms on average, polydispersity index (M_w/M_n) = 1.05–1.10) were separated from a suitable ring–chain equilibrate. Samples crystallized rapidly from the melt were used to determine melting temperature (T_m) as a function of ring size (Clarson *et al.* 1985). The fact that they melted below room temperature limited detailed exploration. The results are shown in Fig. 2 as a plot of T_m against reciprocal ring size (n) in skeletal atoms. Extrapolation to infinite ring size indicates a value of T_m^0 of $\sim -26\,°C$, where T_m^0 would be the equilibrium melting temperature of an infinitely-thick perfect PDMS crystal. The turnover in T_m versus $1/n$ for the larger rings indicates the onset of ring folding, presumably of the larger rings in the distributions.

2.2 Cyclic Poly(oxyethylene)

Cyclic POEs were prepared from commercial narrow-distribution polyethylene glycols (Yu *et al.* 1996). The crystallinities of samples in the range $M_n = 1000$–$10\,000\,\mathrm{gmol}^{-1}$ (average 68–680 skeletal atoms, $M_w/M_n = 1.05$–1.10) were studied (Cooke *et al.* 1998). All showed the WAXS pattern expected for POE. The smaller rings (410 skeletal atoms on average) crystallized into lamellar stacks with spacings (SAXS) and Raman frequencies which corresponded to the extended ring. The largest ring (680 skeletal atoms on average) could be crystallized into a folded conformation as shown in Fig. 3.

Melting temperatures from DSC are plotted in Fig. 2, and show a similar pattern to the PDMS fractions. In both systems ring folding started at (average) 150–200 skeletal atoms. Extrapolation leads to $T_m^0 \sim 70\,°C$, as established for linear POE, while a similar plot for the enthalpy of fusion led to $\Delta_{fus}H^0 \sim 200\,\mathrm{Jg}^{-1}$, also in line with expectation.

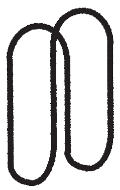


Figure 3
Schematic diagram showing the crystallization of large cyclic poly(oxyethlene) rings (680 skeletal atoms on average) in a folded conformation.

3. Summary

Structural and thermodynamic considerations indicate that cyclic polymers of high molar mass crystallize in the same way as their corresponding linear polymers. Different properties are found for low-molar-mass cyclic polymers, but these can be understood in the light of present knowledge of the crystallinity of linear polymers. Experimental studies on uniform cyclic oligomers yield results that give insight into the nature and energetics of chain folding.

Bibliography

Baxter I, Ben-Haida A, Colquhoun H M, Hodge P, Kohnke F H, Williams D J 1998 Cyclodepolymerization of bisphenol A polysulphone. *Chem. Commun.* 2213–4

Clarson S J, Dodgson K, Semlyen J A 1985 Studies of cyclic and linear poly(dimethylsiloxane)s: 19. Glass transition temperatures and crystallization behaviour. *Polymer* **26**, 930–4

Cooke J, Viras K, Yu G-E, Sun T, Yonemitsu T, Ryan A J, Price C, Booth C 1998 Large cyclic poly(oxyethylene)s: chain folding in the crystalline state studied by Raman spectroscopy, x-ray scattering and differential scanning calorimetry. *Macromolecules* **31**, 3030–9

Lee K-S, Wegner G 1985 Linear and cyclic alkanes (C_nH_{2n+2}, C_nH_{2n}) with n > 100. Synthesis and evidence for chain folding. *Makromol. Chem. Rapid Commun.* **6**, 203–8

Lieser G, Lee K-S, Wegner G 1988 Packing of long-chain cycloalkanes in various crystalline modifications: an electron diffraction investigation. *Colloid Polym. Sci.* **266**, 419–28

Semlyen J A (ed.) 1985 *Cyclic Polymers*. Elsevier, London

Semlyen J A (ed.) 2000 *Cyclic Polymers* 2nd edn. Kluwer, Amsterdam

Yang Z, Cooke J, Viras K, Gorry P A, Ryan A J, Booth C 1997 Chain folding in oligo(oxyethylene)s. Crystallinity of large unsubstituted crown ethers by x-ray scattering and differential scanning calorimetry. *J. Chem. Soc., Faraday Trans.* **93**, 4033–9

Yu G-E, Sinnathamby P, Price C, Booth C 1996 Preparation of large cyclic poly(oxyethylene)s. *Chem. Commun.* 31–2

C. Booth
University of Manchester, UK

Cycloaliphatic Polymers

The polymerization of strained cyclic olefin monomers, such as norbornene, finds its roots in the 1950s with the work of Andersen and Merckling who discovered polymers produced via the ring-opening metathesis polymerization (ROMP) of norbornene. If we ignore carbocationic and free radical initiated reactions (which are known to yield only low oligomers), there are three different mechanisms by which norbornene can be polymerized (Fig. 1). As well as ROMP, there is the "addition" polymerization of norbornenes. The copolymerization of norbornenes with α-olefins forms the basis of ethylene propylene-diene monomer (EPDM) technology, although more recent work is aimed at generating high T_g polymers.

Since 1960, there have been several reports in the literature regarding the addition polymerization of monocyclic olefins such as cyclobutene and cyclopentene or bicyclic olefins such as norbornene (Gaylord *et al.* 1977) using catalysts related to the original Ziegler-Natta inventions. More recently, interest in norbornene and its derivatives (illustrated in Fig. 2) has undergone a renaissance of sorts. This is due to two factors: the development of new, single-site catalysts that will polymerize these types of monomers with good activity, and the realization that polymers containing significant quantities of these bicyclic monomers exhibit unique properties such as high glass-transition temperatures and transparency.

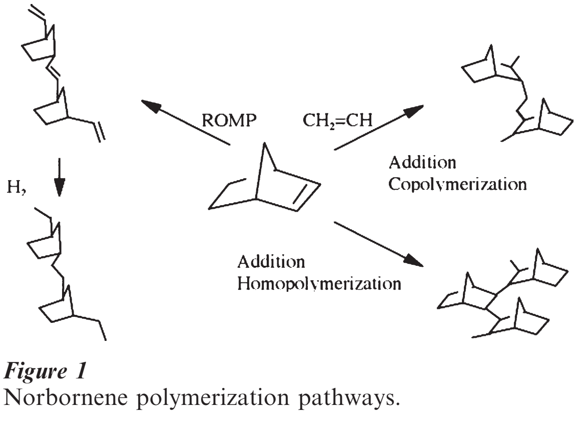


Figure 1
Norbornene polymerization pathways.

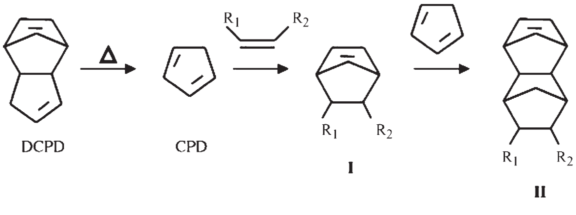


Figure 2
Norbornene monomer synthesis.

Early work at Mitsui Petrochemicals concentrated on copolymerization of the multicyclic olefin, dimethanooctahydronapthalene (DMON, structure II ($R_1 = R_2 = H$) in Fig. 2), using EPDM-type, soluble vanadium catalysts that eventually led to the commercialization of Apel polyolefins (Parshall and Ittel 1992). Later, the utility of metallocene catalysts for cyclic olefin copolymerization was recognized by both Mitsui and Hoechst (Cherdron *et al.* 1994). This led to the joint development of the Topas line of polyolefins, now being marketed by Ticona.

This article will deal exclusively with the amorphous polyolefins generated by addition homopolymerization and ring-opening metathesis polymerization routes and the catalyst technology used in their synthesis. The addition homopolymerization of norbornene was first mentioned in the early 1960s (Sartori *et al.* 1963), with more recent publications (Sen and Lai 1982, Mehler and Risse 1991, Seehof *et al.* 1992) disclosing the use of homogeneous palladium catalysts for the "living" polymerization of norbornene.

There are four families of catalysts that have been reported to catalyze the addition, or "vinyl-type," homopolymerization of norbornene resulting in poly(2,3-bicyclo[2,2,1]hept-2-ene). These four catalyst types are the classical $TiCl_4$-based Ziegler systems (Sartori *et al.* 1963), the zirconocene/aluminoxane (Kaminsky *et al.* 1991, Kaminsky *et al.* 1990) systems, certain electrophilic palladium(II) complexes (Sen and Lai 1982, Mehler and Risse 1991), and well-defined single-component nickel and palladium catalysts developed by B. F. Goodrich.

(i) The titanium catalysts afford only very low molecular-weight materials (molecular weights < 1000) at low yields.

(ii) Zirconocenes were reported to exhibit low activity producing high polymer that decomposes in air at high temperatures before melting and is insoluble in organic solvents.

(iii) Cationic palladium complexes such as $[Pd(CH_3CN)_4][BF_4]_2$, in solvents such as nitromethane, were found to homopolymerize norbornene to high polymers with reasonable solubility in solvents such as tetrachloroethylene, chlorobenzene, and *o*-dichlorobenzene. These polymers have been described with molecular weights in excess of 100 000; indeed the polymerizations have been described as being living in character. These polymers were also found to have high glass transition temperatures ($T_g > 300\,^{\circ}C$).

(iv) Following pioneering work by Sen and Lai (1982) and Risse (Mehler and Risse 1991) in the 1980s, B. F. Goodrich launched a new family of amorphous norbornene-based polymers aimed at the microelectronics industry. These new polymers were made possible by a breakthrough in the area of single-component catalysts based on group VIII transition metals (Goodall *et al.* 1997) These catalysts are characterized by their ability to:

(i) Produce ultra-high glass transition-temperature norbornene homopolymers ($T_g = 355$–$390\,^{\circ}\mathrm{C}$) that can be tailored by copolymerization with 5-alkylnorbornenes.

(ii) Produce polymers with molecular weights exceeding 2000 000 down to oligomers by a novel chain-transfer mechanism with α-olefins.

(iii) Polymerize and copolymerize norbornenes bearing functional groups (R_1 and R_2 in Fig. 2) (carboxylic acid esters, ethers, anhydrides, etc.).

The three families of Goodrich polymers are aimed at a number of microelectronic applications and are high-priced specialties (up to \$3000 per pound (\$6600 per kilogram)).

Each of these polymers exploit the ability of the group VIII metal catalysts to tolerate functional groups and to copolymerize norbornene monomers bearing esters, etc., into polymer backbone. In the case of dielectric polymers (Goodrich 1999, 2000), low levels (2–10 mol %) of triethoxysilylnorbornene (I, $R_1 = H$, $R_2 = Si(OEt)_3$) are used to impart good adhesive properties, the remaining >90% of the monomer being an alkylnorbornene (I, $R_1 = H$, $R_2 =$ alkyl). The alkylnorbornene is selected to tailor the glass transition temperature of the polymer and to impart toughness into the product. The influence of alkyl-substituents on T_g is illustrated in Fig 3.

The optical polymers have a similar composition, primarily an alkylnorbornene to impart toughness and a low (typically <10%) amount of a norbornene bearing an oxygen-containing functional group (to increase chain–chain interactions, improving overall polymer properties).

The use of these polymers in photolithographic applications is possible because of their transparency in the deep UV region (193 nm), and the ability of the nickel and palladium catalysts to polymerize norbornenes bearing a variety of functional groups. However, simply being transparent at the wavelength of interest is only the first hurdle in choosing a candidate polymer backbone. A commercial photoresist must not only undergo the appropriate solubility switch upon irradiation, but the insoluble polymer that remains after aqueous base development must "resist" decomposition under plasma-etching conditions, i. e., conditions in which the exposed portions of the silicon wafer are etched away. Otherwise, 3-D features, and therefore components, could not be built into the wafer. Thus etch rate is a very important parameter that must be measured for candidate photoresist polymers; the lower the etch rate, the better. The cycloaliphatic backbone of the polynorbornenes imparts both transparency and a very low etch rate (i.e., a high reactive ion etch resistance (RIE)). Finally, every monomer unit bears a functional group (typically esters, ethers, and the like) to tailor key parameters such as adhesion and "wettability" (hydrophilicity), T_g, physical properties, and solubility. Finally, a significant portion of the functional groups (typically in the range 10–40%) represent the "solubility switch." A chemically amplified resist system typically contains a photosensitive acid generator (or PAG) such as triarylsulfonium hexafluorophosphate and a polymer containing an acid-sensitive moiety. This moiety or "solubility switch" undergoes an acid (released by photolysis of the PAG) catalyzed deprotection forming a functionality that renders the formerly solvent soluble polymer aqueous base soluble. In practice, derivatives of poly(norbornene) are protected from unexposed dissolution through a judicious choice of an acid-reactive group such as a *t*-butylcarboxylic acid ester (Fig. 4).

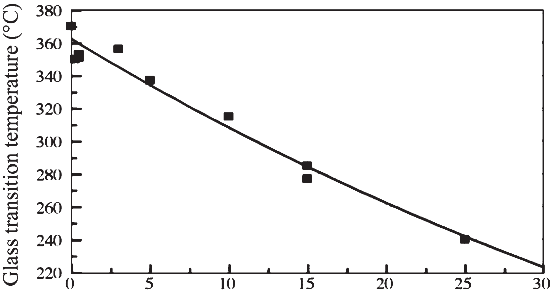


Figure 3
Effect of polymer composition on glass transition temperature.

The catalysts used are all homogeneous nickel or palladium species, usually cationic species with weakly coordinating counter-ions. Although later patents (Goodrich 1996, 1998) describe a broad family of multi-component (or Ziegler-type) catalysts, the first catalysts (Goodrich 1995) were well-defined single-component catalysts of the type illustrated in Fig. 5. A recent patent (Goodrich 2000) has shown that applying a strongly donating phosphine ligand (e.g., tricyclohexyl-phosphine) can enhance the sluggish performance of the palladium catalysts, including both well-defined cationic species and multi-component catalyst systems. Importantly, the phosphine

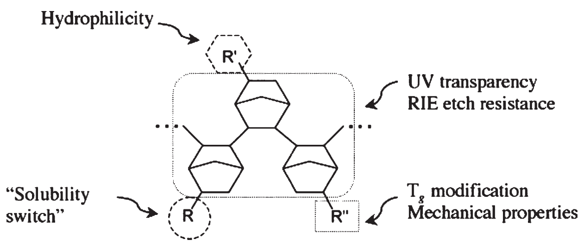


Figure 4
A cycloaliphatic photoresist polymer.

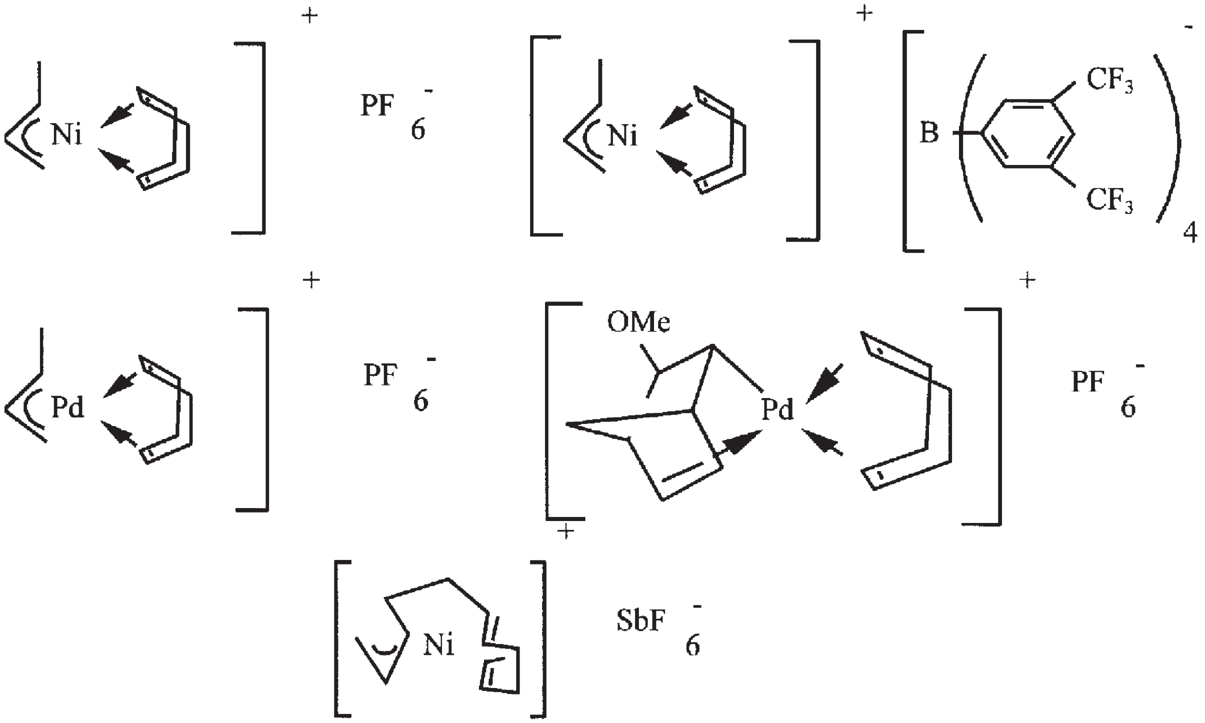


Figure 5
Goodrich's "Naked Metal" catalysts.

ligand greatly enhances the thermal stability of the palladium catalysts, allowing them to be used at elevated temperatures. Overall, the performance of the resulting palladium catalysts surpasses that of the best nickel catalysts by about two orders of magnitude, allowing conversion of up to a million moles of norbornenes per mole of palladium.

In the case of the polymers for photoresist applications there is an extreme need for tolerance of functional groups since, as outlined above, every monomer unit contains a functional group. Such extreme conditions dictate the use of a palladium catalyst, if a cationic catalyst is to be employed. However, neutral nickel catalysts were also found to have a high tolerance of functional groups (presumably because of the diminished electrophilicity of the nickel center compared to cationic counterparts (Boffa and Novak 2000). Thus, in addition to cationic palladium catalysts, certain neutral nickel species are also highly effective systems (Goodrich 1997, 1999), in particular species of the type (L)Ni(Ar^F)$_2$ (where L = covalent ligand, Ar^F = fluorinated aryl group). Examples of such initiators include (toluene)Ni(C_6F_5)$_2$ (first reported as an ethylene dimerization catalyst by Klabunde 1980, Choe *et al.* 1989) and (dimethoxyethane)Ni(3,5-(CF_3)$_2$$C_6H_3$)$_2$. Presumably one of the fluorinated aryl groups serves to tune the electrophilicity of the nickel center while the second inserts a norbornene monomer and becomes a polymer head group.

Other amorphous polyolefins are available from Nippon Zeon (Zeonex) and JSR (Arton). These polymers are based on a two-step process (ring-opening metathesis polymerization (ROMP), followed by hydrogenation). These hydrogenated ROMP copolymers exhibit high glass-transition temperatures (up to about 180 °C) and, like their addition polymer counterparts, find use in optical and electronic applications. Polymerizing alkyl- or ethenyl-substituted DMON monomers (Nippon Zeon 1994, 1996) generates the Nippon Zeon polymers. The polymers, made using two component homogeneous catalysts (typically $MoCl_5$ in combination with an alkylaluminum halide), are hydrogenated employing palladium catalysts. The resulting polymers show excellent transparency and find a number of applications, taking advantage of their optical properties (e.g., optical data storage) and high glass-transition temperature (e.g., medical applications such as syringes where sterilizability is paramount).

Japan Synthetic Rubber with their Arton polymer targets similar applications. In this case the monomer used is typically 8-methyl-8-methoxycarbonyl DMON (the higher Diels-Alder adduct of methylmethacrylate and DCPD), and the ROMP catalyst used is a homogeneous tungsten system (Japan Synthetic Rubber 1992, 1993, 1996).

Bibliography

Boffa L S, Novak B M 2000 Copolymerization of polar monomers with olefins using transition-metal complexes. *Chem. Rev.* **100**, 1479–93

Cherdron H, Brekner M -J, Osan F 1994 Cycloolefin coploymer: a new class of transparent thermoplastics. *Makromol. Chem.* **223**, 121–33

Choe S B, Kanai H, Klabunde K 1989 Catalytic dimerization of ethylene and propylene by (η^6-arene)NiR2 and in combination with alkylaluminum halides. Production of highly active homogeneous catalysts. *J. Am. Chem. Soc.* **111**, 2875–82

Gaylord N G, Deshpande A B, Mandal B M, Martan M J 1977 2,3- and 2,7-bicyclo[2.1.1]hept-2-enes. Preparation and structures of polynorbornenes. *Macromol. Sci. Chem.* **A11**, 1053–70

Goodall B L, Barnes D A, Benedikt G M, Jayaraman S, McIntosh L H, Rhodes L F, Shick R A 1998 Novel heat resistant cyclic olefin polymers made using single component nickel and palladium catalysts: synthesis and applications. *Polym. Prepr. (Am. Chem. Soc., Div. Polym. Chem.)* **39** (1), 216–7

Goodall B L, Barnes D A, Benedikt G M, McIntosh L H, Rhodes L F 1997 *Proceedings of MetCon '97 (Worldwide Metallocene Conference).* Houston, TX, Catalyst Consultants Inc.

Goodrich B F 2000 PCT Application WO 0020472

Goodrich B F 1997, 1999 PCT Applications WO 9733198, WO 9914256, WO 9914635

Goodrich B F 1995 US Patent 5 468 819

Goodrich B F 1999 US Patent 5 929 181

Goodrich B F 1996, 1998 US Patents 5 569730, 5 571 881, 5 741869

Goodrich B F 1999, 2000 US Patents 5 912 313, 6 031 058

Japan Synthetic Rubber 1992, 1993, 1996 US Patents 5 164 469, 5 202 388, 5 518 843

Kaminsky W, Arndt M, Bark A 1991 New results of the polymerization of olefins with metallocene/aluminoxane catalysts. *Polym. Prep. (Am. Chem. Soc., Div. Polym. Chem.)* **32** (1), 467–8

Kaminsky W, Bark A 1992 Copolymerization of ethene and dimethanooctahydronaphthalene with aluminoxane. *Polym. Int.* **28**, 251–3

Kaminsky W, Bark A, Arndt M 1991 New polymers by homogeneous zirconocene/aluminoxane catalysts. *Makromol. Chem., Macromol. Symp.* **47**, 83–93

Kaminsky W, Bark A, Däke I 1990 Polymerization of cyclic olefins with homogeneous catalysts. *Stud. Surf. Sci. Catal.* **56**, 425–38

Kaminsky W, Noll A 1993 Copolymerization of norbornene and ethene with homogeneous zirconocenes/methyl aluminoxane catalysts. *Polym. Bull.* **31**, 175–82

Kaminsky W 1991 Polymerization and copolymerization of olefins with metallocene/aluminoxane catalysts. *Shokubai.* **33**(8), 536–44

Keim W 1990 Nickel—an element with wide application in industrial homogeneous catalysts. *Angew Chem. Int. Ed. Engl.* **29**, 235–44

Klabunde K 1980 π-Arene complexes of nickel(II). Synthesis (from metal atoms) of (π-arene)bis(pentafluorophenyl) nickel(II). Properties, *p*-arene lability, and chemistry. *J. Am. Chem. Soc.* **102**, 4959–66

Klabunde U, Ittel S D 1987 Nickel catalysis for ethylene homo- and copolymerization. *J. Mol. Catal.* **41**, 123–34

Mehler C, Risse W 1991 The palladium(II)-catalyzed polymerization of norbornene. *Makromol. Chem., Rapid Commun.* **12**, 255–9

Nippon Zeon 1994, 1996 US Patents 5 366 812, 5 580 934

Parshall G W, Ittel S D 1992 *Homogeneous Catalysis*, 2nd edn Wiley-Interscience, New York, pp. 63, 224

Sartori G, Ciampelli F, Cameli N 1963 Polymerization of norbornene. *Chim. E. Ind.* **45**, 1478–82

Seehof N, Mehler C, Breunig S, Risse W 1992 Palladium(2+) catalyzed addition polymerizations of norbornene and norbornene derivatives. *J. Mol. Catal.* **76**, 219–28

Sen A, Lai T W 1982 Catalytic polymerization of acetylenes and olefins by tetrakis(acetonitrile)palladium(II). *Organometallics.* **1**, 415–7

Sen A, Lai T W, Thomas R R 1988 Reactions of electrophilic transition metal cations with olefins and small ring compounds. Rearrangements and polymerizations. *J. Organomet. Chem.* **358**, 567–88

Younkin T R, Connor E F, Henderson J I, Friedrich S K, Grubbs R H, Bansleben D A 2000 Neutral, single-component nickel(II) polyolefin catalysts that tolerate heteroataoms. *Science.* **287**, 460–2

B. L. Goodall
Albemarle Corporation, Baton Rouge,
Louisiana, USA

D

Deformation Processing: Texture Evolution

Metals are polycrystalline, and plasticity follows from crystallographic slip processes occuring in individual crystals. In the course of a metal deformation process, this slip activity leads to the rotation of crystal lattices to preferred orientations. The development of preferred orientation in a polycrystalline metal is called *texture*.

The control of texture development in metal deformation processes is often of primary interest in achieving product quality. Modeling of texture evolution thus becomes a tool of great value in gaining fundamental understanding of a particular deformation process. In the following article, several formulations commonly applied for the prediction of deformation texture are outlined. As the final texture of a product often follows from a final, or intermediate, annealing stage, the use of texture transformation models to simulate progression of the deformed texture to a recrystallized texture is also included.

For a more complete treatment involving different metals and deformation paths, including a thorough bibliography, the reader is directed to the work *Texture and Anisotropy* by Kocks *et al.* (1998).

1. *General*

1.1 Crystal Plasticity

(a) Kinematics

So as to draw connections between various modeling approaches in a tractable manner, the discussion is restricted to consideration of a purely viscoplastic response. The deformation **F** of a crystal is then decomposed into a plastic part, $\mathbf{F}^p$, and a rotation, **R***, as

$$\mathbf{F} = \mathbf{R}^* \cdot \mathbf{F}^p \tag{1}$$

The decomposition is shown schematically in Fig. 1. In an individual crystal, the velocity gradient $\mathbf{L}_0^c$ may be related to the shearing rates $\dot{\gamma}^{(s)}$ on slip systems s through:

$$\mathbf{L}_0^c = \dot{\mathbf{F}}^p \cdot \mathbf{F}^{p-1} = \sum_s \dot{\gamma}^{(s)} \mathbf{b}_0^{(s)} \otimes \mathbf{n}_0^{(s)}, \tag{2}$$

where $\mathbf{b}_0^{(s)}$ and $\mathbf{n}_0^{(s)}$ are vectors in the slip direction and normal to the slip plane, respectively. The Schmid tensor, $\mathbf{b}_0^{(s)} \otimes \mathbf{n}_0^{(s)}$, may be divided into the symmetric and skew parts:

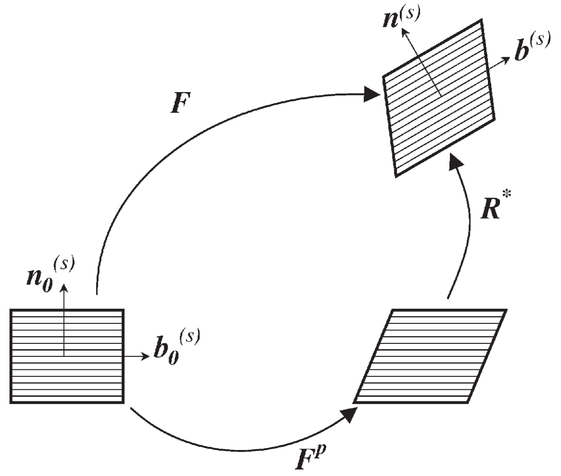


Figure 1
Kinematic decomposition of deformation arising from crystallographic slip.

$$\mathbf{m}_0^{(s)} = \frac{1}{2}\left(\mathbf{b}_0^{(s)} \otimes \mathbf{n}_0^{(s)} + \mathbf{n}_0^{(s)} \otimes \mathbf{b}_0^{(s)}\right) \tag{3}$$

$$\mathbf{q}_0^{(s)} = \frac{1}{2}\left(\mathbf{b}_0^{(s)} \otimes \mathbf{n}_0^{(s)} - \mathbf{n}_0^{(s)} \otimes \mathbf{b}_0^{(s)}\right) \tag{4}$$

respectively. These tensors may also be expressed in the current configuration, and are so indicated in the following by dropping the subscript 0. The collective action of active slip systems in crystal c renders the strain rate $\mathbf{D}^c$ as:

$$\mathbf{D}^c = \sum_s \dot{\gamma}^{(s)} \mathbf{m}^{(s)} \tag{5}$$

and rotation rate $\mathbf{W}^c$ as:

$$\mathbf{W}^c = \sum_s \dot{\gamma}^{(s)} \mathbf{q}^{(s)} \tag{6}$$

The latter term comprises the *plastic spin*.

The evolution of texture is a consequence of the kinematics of slip processes. Let the spin of the grain be denoted by $\mathbf{W}^g$. The *lattice spin* of the crystallographic axes, **W***, is the difference between the spin of the grain axes and the plastic spin:

$$\mathbf{W}^* = \mathbf{W}^g - \mathbf{W}^c \tag{7}$$

The lattice spin is associated with the orientation **R*** as $\mathbf{W}^* = \dot{\mathbf{R}}^* \cdot \mathbf{R}^{*-1}$. Introducing **R*** into Eqn. (7)

provides the evolution equation:

$$\dot{\mathbf{R}}^* = (\mathbf{W}^g - \mathbf{W}^c) \cdot \mathbf{R}^* \tag{8}$$

It is this last relationship that provides for the evolution of the crystallographic orientation of a crystal with deformation.

(b) Relation of crystal deformation to that at the macroscale

At the level of the polycrystal, an aggregate of crystals, the velocity gradient is related to the deformation $\mathbf{F}$ as:

$$\mathbf{L} = \dot{\mathbf{F}} \cdot \mathbf{F}^{-1} \tag{9}$$

The velocity gradient is decomposed into symmetric ($\mathbf{D}$) and skew-symmetric ($\mathbf{W}$) parts as:

$$\mathbf{L} = \mathbf{D} + \mathbf{W} \tag{10}$$

It is these factors:

- the relation of $\mathbf{D}$ to $\mathbf{D}^c$;
- the relation of $\mathbf{W}$ to $\mathbf{W}^g$;
- the constitutive assumptions in the determination of the crystal stress $\boldsymbol{\sigma}^c$;
- equilibrium considerations

that provide differences among models for polycrystal plasticity. The averaging of crystal response to describe the deformation behavior of the polycrystal necessarily addresses the interaction among grains. Models differ in their treatment of compatibility and equilibrium.

(c) Crystal stress

Equilibrium considerations enter into the relationship of crystal response to the macroscale. Slip on a slip system s follows the Schmid law:

$$\mathbf{m}^s : \boldsymbol{\sigma}^c \leqslant \tau^s \tag{11}$$

where $\boldsymbol{\sigma}^c$ is the stress in the crystal c and τ^s denotes the strength of the slip system. Equation (11) represents a system of hyperplanes, and the inner envelope which they define poses a yield surface. Given a prescribed deformation $\mathbf{D}$, the stress must coincide with a vertex of the single crystal yield surface. One may select the stress state through identification of the vertex stress $\boldsymbol{\sigma}$ from a list of possible vertex stress states $\boldsymbol{\sigma}^*$ (presented by the collection of hyperplanes), using the maximum work principle of Bishop and Hill (1951a, 1951b):

$$(\boldsymbol{\sigma} - \boldsymbol{\sigma}^*) : \mathbf{D} \geqslant 0 \tag{12}$$

While the stress state is thus available, given a prescribed deformation, the slip system shear rates are available only when the number of slip systems equals the number of deformation degrees of freedom. This is a rare condition in practice.

To model texture evolution, it is clearly necessary to establish the spin, $\mathbf{W}^c$. This in turn mandates knowledge of the shears $\dot{\gamma}^{(s)}$. In application, the shears follow from stress considerations and require a constitutive assumption. A viscoplastic relationship introduced in Hutchinson (1976) is:

$$\frac{\dot{\gamma}^{(s)}}{\dot{\gamma}_0} = \left|\frac{\tau^{(s)}}{\bar{\tau}}\right| \sin(\tau^{(s)}), \tag{13}$$

where the exponent n provides stress rate sensitivity and $\dot{\gamma}_0$ is a reference strain rate.

Equations (5) and (13) are combined to deliver a relation for the inelastic crystal deformation rate:

$$\boldsymbol{D}^c = P^{(\mathrm{sec})}\colon \boldsymbol{\sigma}'^{\mathrm{c}}, \tag{14}$$

where $\boldsymbol{\sigma}'^{\mathrm{c}}$ is the deviatoric stress and $P^{(\mathrm{sec})}$ is the secant modulus (a fourth-order tensor):

$$P^{(\mathrm{sec})} = \sum_{\kappa} \frac{\dot{\gamma}_0}{\hat{\tau}} \left|\frac{\tau^{(s)}}{\hat{\tau}}\right|^{n-1} \mathbf{m}^{(s)} \otimes \mathbf{m}^{(s)} \tag{15}$$

Note that the Cauchy stress is partitioned in the usual manner: $\boldsymbol{\sigma} = \boldsymbol{\sigma}' - p\mathbf{I}$, where p is the hydrostatic portion of the stress and $\mathbf{I}$ is the identity tensor. Use of Eqn. (14) renders it unnecessary to define vertices of the single crystal yield surface, and removes ambiguity in the selection of active slip systems.

Generally speaking, the simulation of texture evolution for a particular crystallographic orientation proceeds as:

(i) a deformation rate $\mathbf{L}^{\mathrm{c}}$ is derived from the macroscale deformation $\mathbf{L}$;
(ii) $\boldsymbol{\sigma}'^{\mathrm{c}}$ is developed using Eqns. (14) and (15);
(iii) slip system shear stress follows from Eqn. (11), assuming equality;
(iv) slip system shear rates are recovered (Eqn. (13));
(v) the plastic spin is now available from Eqn. (6);
(vi) the orientation is updated using the evolution equation, Eqn. (8).

These steps are performed incrementally, over discrete time steps. The simulation of texture for a deformed metal results from carrying out the above procedure for "many" orientations.

1.2 Description of Texture

The orientation of a crystal is described by the orthogonal tensor, $\mathbf{R}^*$. For the purposes of the present discussion, the evolution of the crystal orientation is achieved by application of Eqn. (8). Texture results from characterization of the distribution of crystal orientations $\mathbf{R}^*$ in a sample. The resulting crystallographic texture, $f(\mathbf{R}^*)$, is given by the orientation distribution function. The ODF can be expressed in terms of a series expansion to (symmetrized) spherical harmonic functions (Bunge 1982):

$$f(\mathbf{R}^*) = \sum_{l=1}^{l_{\max}} C_l^{\mu\nu} \cdot T_l^{\mu\nu}(\mathbf{R}^*), \tag{16}$$

where $C_l^{\mu\nu}$ are the series expansion coefficients and $T_l^{\mu\nu}$ are the spherical harmonics. The orientation of a crystal may also be expressed as a function of three Euler angles:

$$\mathbf{R}^* = \mathbf{R}^*(\psi, \theta, \phi) \tag{17}$$

Many conventions are possible; both the symmetric angle convention due to Kocks (Kocks *et al.* 1998) and a widely used convention developed by Bunge (1982) will be used in the following.

Texture is typically described through a distribution of density in an orientation space spanned by the three Euler angles (Randle and Engler 2000). The orientation distribution, using the Kocks convention, for rolled aluminum is shown in Fig. 2. In these sections from the orientation distribution function, two of the Euler angles, ψ and θ, provide polar coordinates within each quadrant. The quadrants have constant values of the third angle, φ. When using the Kocks convention, the projection of all sections yields the $\langle 100 \rangle$ pole figure (labeled PROJ in Fig. 2).

2. *Models for Texture Development*

A brief survey is made of models for the development of texture. As many of the modeling techniques to be discussed were developed around the rolling process—and to provide a succinct example—we will focus on plane strain compression of f.c.c. metals.

Rolled metal sheet and plate products are an essential component of engineering structures. Application is widespread in the building, transportation, and packaging industries. As a consequence, properties of rolled products are of considerable interest. Anisotropy is one such property, and follows from the directional character of the rolling process. At the centerplane, rolling approximates a state of plane strain compression. Adopting the usual Cartesian reference frame, taking the 3-direction as the compression axis and the 1-direction as the extension axis implies the plane strain deformation:

$$F = \begin{bmatrix} a & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \frac{1}{a} \end{bmatrix} \tag{18}$$

with $a > 1$. This deformation path provides for study using models that predict texture evolution only for a single representative volume element, or material point. Redundant shear is developed away from the centerplane, due to frictional contact with the work rolls.

The quality of rolled products follows (in part) from the ability to understand and control the rolling process. Much effort has been placed into developing predictive models for rolling, including the prediction of texture. We will use the results of modeling efforts to "look inside" the roll bite and examine texture development.

The f.c.c. rolling texture is characterized by development along the β-fiber (Hirsch and Lücke 1988a), with $\langle 110 \rangle$ tilted 60° to the rolling direction. Three particular components distinguish the fiber:

- the copper, or C, component with Miller indices $\{112\}\langle 111 \rangle$;
- the S component, near $\{123\}\langle 634 \rangle$;
- the brass, or B, component with indices $\{011\}\langle 211 \rangle$.

Note that the Miller indices of the form $\{hkl\}\langle uvw \rangle$ refer to the crystallographic plane $\{hkl\}$ coincident with the physical rolling plane sharing a normal to the sheet surface—with crystallographic direction $\langle uvw \rangle$ aligned in the direction of rolling. The β fiber progresses through sections of the ODF from the orientation in the $\varphi = 45°$ section, through the orientation, and terminates at the orientation in

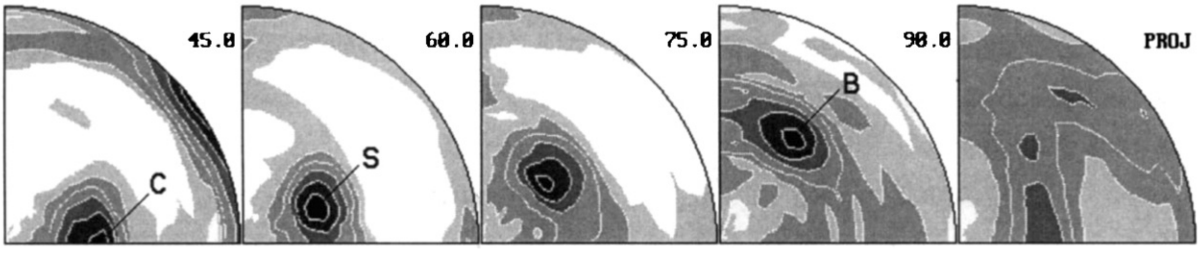


Figure 2
ODF for commercial purity aluminum rolled to 80% reduction. Sections of the ODF containing the β-fiber are shown. Rolling direction is horizontal, transverse direction is vertical and the sheet normal direction is out of the page.

the $\varphi = 90°$ section. The development along a fiber in orientation space traversing these preferred orientations may be seen in Fig. 2.

2.1 Full Constraints (Taylor) Model

Linking the crystal and macroscale (polycrystal) kinematics as:

$$\mathbf{D} = \mathbf{D}^c \tag{19}$$

$$\mathbf{W} = \mathbf{W}^g \tag{20}$$

renders the *full constraints* model (Hirsch and Lücke 1988b). That is, the deformation rate adopted in crystals comprising the polycrystal is identical to that at the macroscale. Equilibrium conditions among crystals are not satisfied. The predicted texture for 80% reduction is shown in Fig. 3(a). This texture is distiguished by the Taylor component, {4 4 11}⟨11 11 8⟩, and may be seen in the $\varphi = 45°$ section. The peak in this section is slightly displaced toward the origin, as compared to the experimental texture.

2.2 Relaxed Constraints Model

In the limit as $a \to \infty$, the interaction between two grains "A" and "B" would be limited to a single plane interface with the normal in the 3-direction. In this limiting condition, equilibrium along the interface results in stress uniformity across the grain thickness:

$$\sigma_{33}^{\mathrm{A}} = \sigma_{33}^{\mathrm{B}} \tag{21}$$

$$\sigma_{31}^{\mathrm{A}} = \sigma_{31}^{\mathrm{B}} \tag{22}$$

$$\sigma_{23}^{\mathrm{A}} = \sigma_{23}^{\mathrm{B}} \tag{23}$$

and compatibility is limited to distortion within the plane:

$$D_{11}^{\mathrm{A}} = D_{11}^{\mathrm{B}} \tag{24}$$

$$D_{22}^{\mathrm{A}} = D_{22}^{\mathrm{B}} \tag{25}$$

$$D_{12}^{\mathrm{A}} = D_{12}^{\mathrm{B}} \tag{26}$$

Such considerations lead to the *relaxed constraints* model. It is a first step in satisfying equilibrium in the case of grains that are flat, a condition approached in rolling to large reductions. The key feature is an improved simulation of rolling textures, particularly at high strains (Fig. 3(b)).

2.3 Viscoplastic Self-consistent Model

The viscoplastic self-consistent (VPSC) model introduces a grain as a viscoplastic inclusion, or inhomogeneity, in a homogeneous medium. Loading of surface Γ_V of the volume V leads to a non-uniform stress field in the vicinity of the inclusion. The stress field is given by:

$$\boldsymbol{\sigma}'(\mathbf{x}) = [P^{(\mathrm{tg})}]^{-1} : [\mathbf{D}(\mathbf{x}) - \mathbf{D}^{c0}] \quad \text{in} \quad \Omega \tag{27}$$

$$\boldsymbol{\sigma}'(\mathbf{x}) = [P^{(\mathrm{tg})}]^{-1} : [\mathbf{D}(\mathbf{x}) - \mathbf{D}^{0}] \quad \text{in} \quad V - \Omega \tag{28}$$

$$\boldsymbol{\sigma}'(\mathbf{x}) = [P^{(\mathrm{tg})}]^{-1} : [\mathbf{D} - \mathbf{D}^{0}] \quad \text{in} \quad \Gamma_V \tag{29}$$

and

$$\nabla \cdot \boldsymbol{\sigma} = 0 (\text{equilibrium}) \tag{30}$$

where the tangent modulus is given by:

$$P^{(\mathrm{tg})} = nP^{(\mathrm{sec})} \tag{31}$$

The terms $\mathbf{D}^{c0}$ and $\mathbf{D}^0$ are reference deformation rates, following from expansion of Eqn. (14) about a reference stress $\boldsymbol{\sigma}'^0$. The problem of a grain (Ω) contained in a homogeneous effective medium (V–Ω, hereafter referred to as the HEM) is embodied by Eqns. (27) to (30). For details on solution, the reader should consult Kocks *et al.* (1998, Chap. 11).

To model the response of a polycrystal, each grain is considered as an inclusion in the HEM. The linearization of the boundary value problem guarantees that stress and strain rate are uniform within the domain of the inclusion. The HEM represents the average response of all grains in the aggregate. This is achieved by an implicit scheme in which the viscoplastic response, $P^{(\mathrm{sec})}$, of the HEM is derived from the response of individual grains. Further, the average stress of all grains is constrained to be equivalent to the macroscale stress:

$$\sum_c w_c \sigma^c = \sigma \tag{32}$$

where w_c is a weighting factor for grain (or crystal) c. Stated another way, the viscoplastic response of the HEM and the corresponding crystal averages are *self-consistent*.

For cubic metals, the predicted textures are similar in character to those developed using the above approaches of full and relaxed constraints. However, there exists an interplay between the grain–matrix interaction and material anisotropy that does not exist in these above models. For example, the location and density of the C component is dependent on the choice of a tangent or secant stiffness. This follows from deviations in shear deformations among grains for the different modeling choices. An example of texture development using the VPSC procedure is given in Fig. 3(c). Note the enhanced development of the (brass) component, relative to the full and relaxed constraints models. This is due to development of shearing in the rolling plane (Hirsch and Lücke

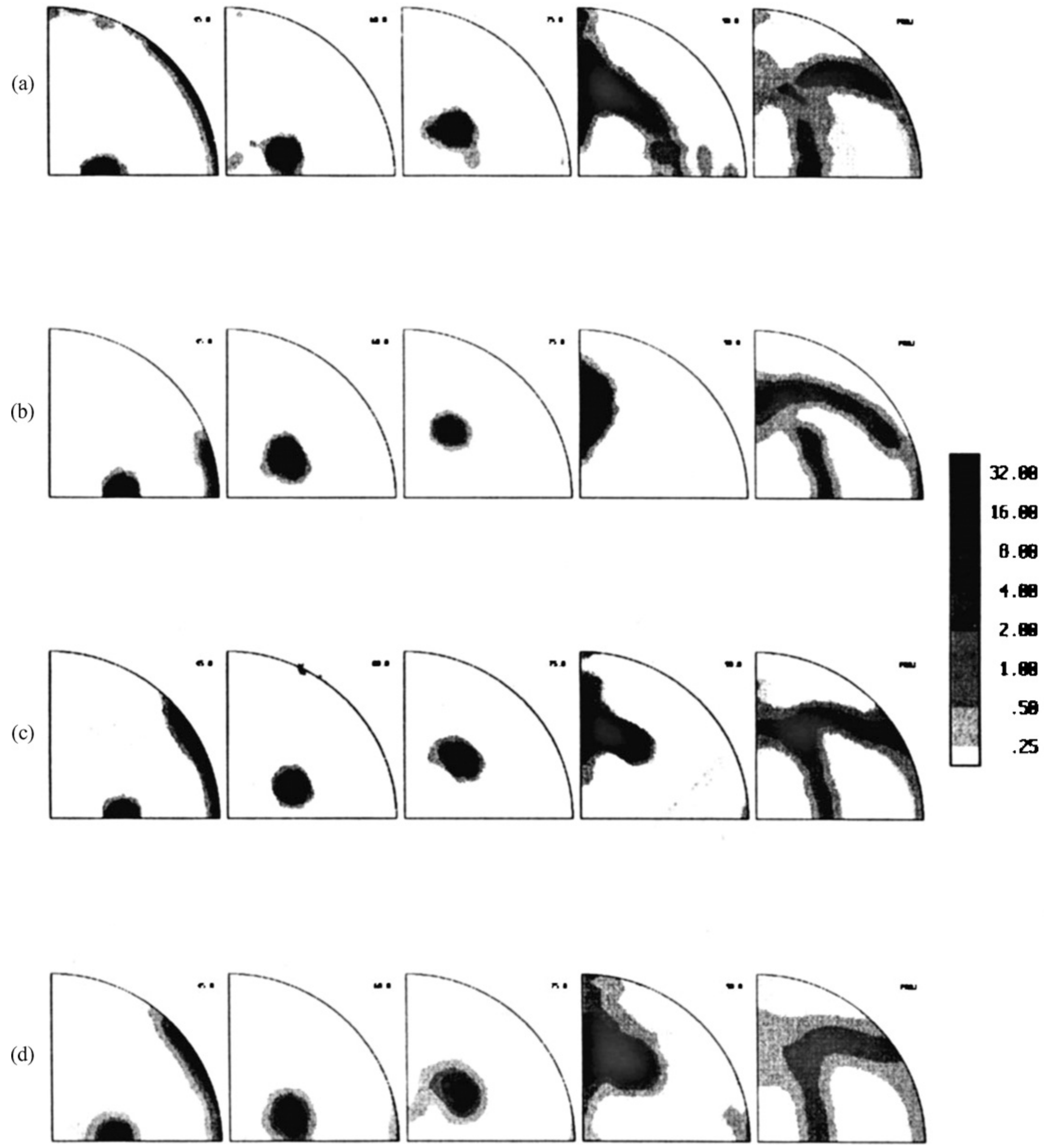


Figure 3
Predicted ODFs for plane strain compression of an f.c.c. metal to 80% reduction, developed using simulation procedures discussed in text: (a) full constraints, (b) relaxed constraints, (c) viscoplastic self-control, (d) finite element polycrystal.

1988a), and is both in better agreement with experiment and an expected consequence of a grain interacting with its environment.

The VPSC approach finds real strength and application in very anisotropic materials, in particular those materials with few available slip systems. Here, grain-matrix interaction becomes a necessity, as one cannot explicitly relate the response of crystals to the macroscopic deformation.

2.4 Finite Element Models

The interaction among grains comprising the polycrystal may be addressed by posing a boundary value problem for the aggregate. The finite element method offers a solution procedure for the boundary value problem on a body Ω, as governed by the field equations:

$$\nabla \cdot \sigma = 0 \quad \text{in} \quad \Omega$$

$$\sigma \cdot \mathbf{n} = \hat{\mathbf{t}} \quad \text{on} \quad \Gamma_\sigma$$

$$\mathbf{u} = \hat{\mathbf{u}} \quad \text{on} \quad \Gamma_\mathbf{u} \tag{33}$$

where $\boldsymbol{\sigma}$ is the Cauchy stress, $\mathbf{n}$ is the normal to the surface Γ_σ having tractions $\hat{\mathbf{t}}$, and $\hat{\mathbf{u}}$ provides the velocity boundary condition on surface $\Gamma_\mathbf{u}$. Introducing a weighted residual built upon the equilibrium equation:

$$\int_B \sigma' : \nabla \mathbf{v} \mathrm{d}V - \int_B p \nabla \cdot \mathbf{v} \mathrm{d}V = \int_{B_\sigma} \hat{\mathbf{t}} \cdot \mathbf{v} \mathrm{d}S \tag{34}$$

where $\mathbf{v}$ are weighting functions. Constitutive response is introduced using Eqn. (14). The usual velocity–pressure formulation follows from specification of a constraint for incompressibility.

The polycrystal is spatially discretized into component crystals, with crystals discretized using one or more finite elements. Boundary conditions are prescribed such that the aggregate response may be linked to a homogeneous deformation characteristic of a material point (inhomogeneity resides within the aggregate following from grain interaction). A section through a sample mesh for the deformation of an f.c.c. polycrystal is shown in Fig. 4.

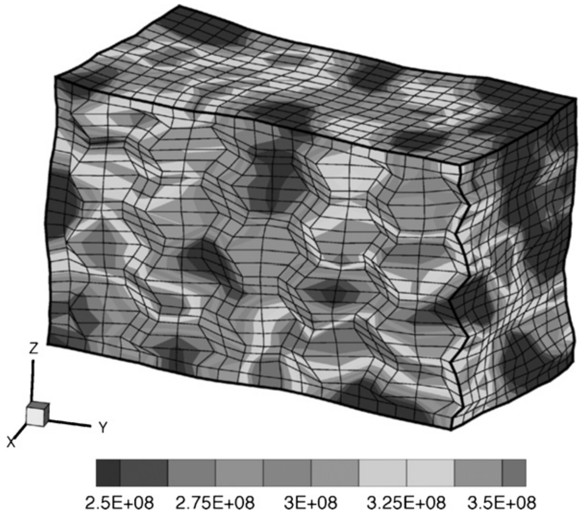


Figure 4
Development of deviatoric stress in model of an f.c.c. polycrystal (shown in Pa). Individual grains are idealized as Wigner–Seitz cells.

Texture developed using a finite element model is given in Fig. 3(d). As compared to the *full constraints* and *relaxed constraints* model, texture tends to be a bit more diffuse. This is a result of local shear deformations, such inhomogeneity arising from grain-to-grain interactions. A distinct peak at the (brass) orientation is predicted—similar to the experimental result—and is a result of the shearing in the rolling plane. Both the lessening of overall texture intensity and the prediction of the brass component are features desirable in modeling the experimental plane strain response of f.c.c. metals.

3. Prediction of Recrystallization Texture from the Deformed Texture

Recrystallization of deformed metallic materials proceeds by the formation of new, undeformed grains in the as-deformed microstructure, their so-called "nucleation," and their subsequent growth into the deformed neighborhood. As these mechanisms are characterized respectively by the formation and the motion of high-angle grain boundaries, recrystallization generally leads to a change in the as-deformed crystallographic texture. These textural changes are often interpreted in terms of one of two theories. (i) In the case of oriented nucleation, it is assumed that the preferred formation of special orientations determines the final recrystallization texture; (ii) in the case of oriented growth, it is assumed that, starting from a broad spectrum of nucleus orientations, those with the best growth conditions with respect to the deformed matrix grow fastest and, thus, dominate the recrystallization texture.

In many f.c.c. materials, including copper and aluminum alloys, it has often been shown that the orientations obtained after recrystallization can be related to the rolling texture through a 40° $\langle 111 \rangle$ rotation (Lücke 1974). Detailed local texture analysis of the newly forming grains indeed substantiates a preference of grains with an approximate 40° $\langle 111 \rangle$ orientation relationship to the rolling texture (Engler 1996).

If the recrystallization of a sample is totally governed by grains with a 40° $\langle 111 \rangle$ orientation relationship to the deformation texture, the resulting recrystallization textures can be simulated by a rotation of the deformation texture by $\pm 40°$ around all possible $\langle 111 \rangle$-axes. However, the above-mentioned "nucleation" and "growth" theories for recrystallization texture interpretation represent the two limiting cases. In most cases, a simulation of recrystallization textures solely based on either of these two theories fails. Instead, a combination of both theories, in the sense of a growth selection out of a limited spectrum of preferentially formed nucleus orientations, has been shown to account for the recrystallization textures of most aluminum alloys (e.g., Engler 1996). A reasonable modeling strategy is to multiply the probability function $f^{\mathrm{nucl}}(\mathbf{R}^*)$ of the occurrence of the nucleus orientations, and the probability function $f^{\mathrm{grow}}(\mathbf{R}^*)$ of their growth, to simulate the probability of the orientations likely to arise in the final recrystallization texture $f^{\mathrm{sim}}(\mathbf{R}^*)$:

$$f^{\mathrm{sim}}(\mathbf{R}^*) = f^{\mathrm{nucl}}(\mathbf{R}^*) \cdot f^{\mathrm{grow}}(\mathbf{R}^*) \tag{35}$$

3.1 Modeling Recrystallization Textures Through Texture Transformations

With a view to the preferred growth of grains with a 40° $\langle 111 \rangle$ orientation relationship to the deformed matrix, the growth probability $f^{\mathrm{grow}}(\mathbf{R}^*)$ is assumed to correspond to the 40° $\langle 111 \rangle$-transformed rolling texture. Thus, the rolling texture $f^{\mathrm{roll}}(\mathbf{R}^*)$—be it experimental or a texture developed using the above simulation procedures for deformation—has to be transformed according to:

$$f^{\mathrm{trans}}(\mathbf{R}^*) = \frac{1}{8} \sum_{i=1}^{8} f^{\mathrm{roll}}(\Delta \mathbf{R}^* \cdot \mathbf{R}^*), \tag{36}$$

where the $\Delta \mathbf{R}_i^*$ denote the eight different 40° $\langle 111 \rangle$-rotations.

From the C-coefficients of the rolling texture, Eqn. (16), the $C^{\mu' v}$-coefficients of the transformation texture can be calculated using the following transformation law (Bunge 1982):

$$C^{\mu' v} = \frac{1}{2l+1} \sum_{\mu=1}^{M(l)} C_l^{\mu v} \cdot W_l^{\mu' \mu}, \tag{37}$$

where W is the transformation law function. If W consists only of one orientation relationship $\Delta \mathbf{R}^*$ including its symmetrical variants, then W is defined as follows:

$$W_l^{\mu' \mu} = (2l+1) T_l^{\mu' \mu}(\Delta \mathbf{R}^*) \tag{38}$$

3.2 Recrystallization Texture of Cold-rolled Aluminum Alloys

Aluminum alloys are characterized by a high stacking fault energy; hence, nucleation through twinning does not play a significant role. Rather, nucleation takes place by enhanced subgrain growth in the vicinity of structural heterogeneities ("preexisting nuclei"), where substantially larger local misorientations exist than in the homogeneously deformed matrix (Doherty 1978). In commercially rolled aluminum alloys, three nucleation sites are of main importance, namely: cube-bands, grain boundaries, and large second-phase particles (Juul Jensen *et al.* 1985, Hjelen *et al.* 1991, Weiland *et al.* 1994, Engler 1996). The final recrystallization textures then result from a competition between the various texture components in dependence on the number and efficiency of the corresponding nucleation sites. For a simulation of the nucleation texture, the contribution of the three nucleation sites may be superimposed according to:

$$f(\mathbf{R}^*)^{\mathrm{nucl}} = x_{\mathrm{Cube}} \cdot f(\mathbf{R}^*)^{\mathrm{nucl}}_{\mathrm{Cube}} + x_{\mathrm{GB}} \cdot f(\mathbf{R}^*)^{\mathrm{nucl}}_{\mathrm{GB}} + x_{\mathrm{PSN}} \cdot f(\mathbf{R}^*)^{\mathrm{nucl}}_{\mathrm{PSN}} \tag{39}$$

where the probability functions $f(\mathbf{R}^*)^{\mathrm{nucl}}_{\mathrm{Cube}}$, $f(\mathbf{R}^*)^{\mathrm{nucl}}_{\mathrm{GB}}$, and $f(\mathbf{R}^*)^{\mathrm{nucl}}_{\mathrm{PSN}}$ denote the orientation spectra of the cube-bands, grain boundaries, and particles, respectively. The weight factors x_i denote the corresponding probability or efficiency of nucleation (with $x_{\mathrm{Cube}} + x_{\mathrm{GB}} + x_{\mathrm{PSN}} = 1$).

The actual distribution of nucleus orientations may be determined by electron microscope-based local texture analysis, using electron back-scattered diffraction (EBSD) in an SEM or microdiffraction in a TEM. From the resulting microtexture data, the distribution of the nucleus orientations, i.e., the nucleation probability function $f^{\mathrm{nucl}}(\mathbf{R}^*)$, can be computed by associating each single orientation with a Gauss-type peak with a given scatter width in orientation space (Bunge 1982). As an alternative procedure, the probability functions may be developed in a heuristic fashion:

- for the nucleation of the cube-orientation, $f(\mathbf{R}^*)^{\mathrm{nucl}}_{\mathrm{Cube}}$, a function was computed by generating an ODF which consists of the cube-orientation and its rotations up to 40° about both the rolling direction and the transverse direction;
- for the nucleation probability at the grain boundaries, $f(\mathbf{R}^*)^{\mathrm{nucl}}_{\mathrm{GB}}$, a weakened rolling texture is adopted, which gives a good approximation of the nucleus orientations;
- for particle-stimulated nucleation, $f(\mathbf{R}^*)^{\mathrm{nucl}}_{\mathrm{PSN}}$, random nucleation was assumed, i.e., set equal to one.

With these three nucleation probabilities, as well as the corresponding transformed rolling texture to consider selected growth, the recrystallization texture of commercial purity aluminum (AA1145) was simulated. The simulation was optimized by fitting the weight factors x_i (see Eqn.(39)) so as to yield best agreement with the experiments. The resulting recrystallization texture, obtained with $x_{\mathrm{Cube}} = 0.2$, $x_{\mathrm{GB}} = 0.7$, and $x_{\mathrm{PSN}} = 0.1$, is given in Fig. 5(a). From comparison with the recrystallization texture obtained experimentally (Fig. 5(b)), the very good resemblance both in the position of the peaks as well as in their intensities is apparent.

The very general character of the above procedure enables the study of different microstructural parameters—including particle distribution, initial texture,

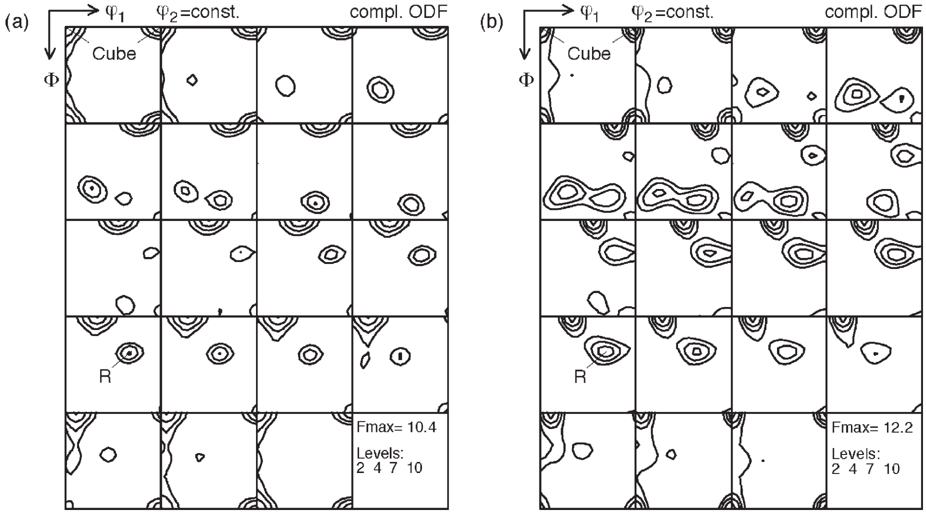


Figure 5
(a) Simulation of the recrystallization texture and (b) experimental recrystallization texture of aluminum alloys, taking all three types of nucleation sites into account (commercial purity Al, 95% cold-rolled, annealed for 1000 s at 350 °C).

and initial grain size—and/or processing parameters—including strain, strain rate, and deformation temperature. Thus, the recrystallization textures of aluminum alloys can be predicted for a wide range of microstructural characteristics and processing parameters.

4. *Conclusion*

With a combination of models to simulate texture changes accompanying deformation and recrystallization, such as the ones introduced in this article, the texture evolution during the entire process chain of sheet metal processing can be simulated from the as-cast ingot to the final annealed sheet. Eventually, this should enable the prediction of property changes caused by changes in the thermomechanical processing, so improving product quality and avoiding laborious and expensive plant trials.

Acknowledgments

The simulation results for the viscoplastic self-consistent model, as well as many valuable suggestions on this entry, were provided by Dr. Carlos Tomé. We also thank Mr. Schalk Kok for his careful review.

Bibliography

Bishop J F W, Hill R 1951a A theory of the plastic distortion of a polycrystalline aggregate under combined stresses. *Phil. Mag.* **42**, 414–27

Bishop J F W, Hill R 1951b A theoretical derivation of the plastic properties of a polycrystalline face-centred metal. *Phil. Mag.* **42**, 1298–307

Bunge H J 1982 *Texture Analysis in Materials Science*. Butterworths, London

Doherty R D 1978 In: Haessner F (ed.) *Recrystallization of Metallic Materials*. Riederer Verlag, pp. 23–61

Engler O 1996 Nucleation and growth during recrystallisation of aluminium alloys investigated by local texture analysis. *Mater. Sci. Tech.* **12**, 859–72

Hirsch J, Lücke K 1988a Mechanism of deformation and development of rolling textures in polycrystalline f.c.c. metals—i. description of rolling texture development in homogeneous Cu-Zn alloys. *Acta Metall.* **36**, 2863–82

Hirsch J, Lücke K 1988b Mechanism of deformation and development of rolling textures in polycrystalline f.c.c. metals—ii. simulation and interpretation of experiments on the basis of Taylor-type theories. *Acta Metall.* **36**, 2883–904

Hjelen J, Ørsund R, Nes E 1991 On the origin of recrystallization textures in aluminum. *Acta Metall. Mater.* **39**, 1377–404

Hutchinson J W 1976 Bounds and self-consistent estimates for creep of polycrystalline materials. *Proc. R. Soc. Lond. A* **348**, 101–27

Juul Jensen D, Hansen N, Humphreys F J 1985 Texture development during recrystallization of aluminium containing large particles. *Acta Metall.* **33**, 2155–62

Kocks U F, Tomé C N, Wenk H-R 1998 *Texture and Anisotropy*. Cambridge University Press, Cambridge

Lücke K 1974 The orientation dependence of grain boundary motion and the formation of recrystallization textures. *Can. Met. Quart.* **13**, 261–74

Randle V, Engler O 2000 *Introduction to Texture Analysis, Macrotexture, Microtexture and Orientation Mapping*. Gordon and Breach, London

Weiland H, Rouns T N, Liu J 1994 The role of particle-stimulated nucleation during recrystallization of an aluminum-manganese alloy. *Z. Metallk.* **85**, 592–7

A. J. Beaudoin
University of Illinois, Urbana, Illinois, USA

O. Engler
VAW Aluminium, Bonn, Germany

Deformation Textures

Virtually all crystalline materials encountered in everyday life are in the form of aggregates of large numbers of individual crystals (or grains) termed polycrystals. The processing of these polycrystals into useful components or structures generally involves one or more, often several, stages of controlled material deformation to form the desired shape. Such deformation processing is often through large strains. The activation of slip (or deformation twinning) systems during deformation of a crystal generally leads to both a change in the crystal shape (strain) and crystal orientation (rotation). For a given applied strain path there are, however, certain crystal orientations for which the combined action of the multiple slip (and/or twinning) systems can accommodate the imposed deformation without causing any change in the crystal orientation. As a result the deformation processing imparts preferred orientations or texture to the polycrystal, as the individual crystals rotate toward the stable orientations.

Texture is of considerable importance because of the marked anisotropy it imparts to a material. Texture in rolled steel and aluminum sheets has a major impact on the mean plastic strain ratio and planar anisotropy which control the limiting draw ratio and extent of earing during subsequent deep drawing operations which are crucial to major manufacturing industries (e.g., beverage cans). Magnetization of b.c.c. Fe–Si alloys is "easy" along $\langle 100\rangle$ direction and "hard" along $\langle 111\rangle$, so that texture plays an large role in reducing power loss through hysteresis in transformer cores made from Fe–Si laminates. Texture is also significant for the semiconductor industry, since controlling the texture of aluminum and copper metallization tracks can give significant improvements in their resistance to electromigration failures.

There have been a great number of investigations into the deformation textures of metals and metallic alloys (Dillamore and Roberts 1965, Bunge 1987), and also those of natural minerals (Wenk 1999), where texture is termed crystal fabric. Traditional texture (macrotexture, see *Texture*) measurement has been carried out using x-ray (Bunge 1987) and neutron diffraction (Bacon 1975). Over recent years the electron back scatter diffraction (EBSD) technique (Wilkinson and Hirsch 1997) has allowed the orientation of individual grains to be measured directly at high spatial resolution allowing textural and microstructural analysis to be linked. This article will briefly outline the well-established experimentally determined fiber textures that result from axisymmetric deformation (e.g., wire drawing, extrusion) and rolling (approximate plane strain deformation), before describing the theoretical frameworks used to model/simulate texture evolution during large strain deformation processing. The most thorough text describing the measurement and simulation of texture and its effects on the anisotropy of physical properties is the book edited by Kocks *et al.* (1998).

1. Axisymmetric Deformation Textures

The simplest axisymmetric deformations are uniaxial tension and uniaxial compression; however, in tension necking prevents the high strains required to develop strong textures. Instead, extrusion and wire drawing can be used to obtain the necessary larger tensile elongations since the constraints on the lateral deformation then prevent necking. In this way the effects of both axial elongations and compressions on texture, have been examined for a range of metals and alloys (Dillamore and Roberts 1965). The symmetry of the deformation requires that only the crystallographic axis along the deformation direction need be given to describe the texture, since in true axisymmetry the orthogonal directions will be randomly distributed. Table 1 shows the commonly found textures for axisymmetric deformation of f.c.c., b.c.c., and h.c.p. metals and alloys. Tensile deformation leads to a $\langle 111\rangle+\langle 100\rangle$ double fiber in the f.c.c. metals. The proportions of the two fibers varies with stacking fault energy (SFE), the $\langle 111\rangle$ fiber making up a large fraction $\sim 90\%$ in aluminum (high SFE) and smaller fraction $\sim 15\%$ in silver (low SFE), with copper and nickel showing intermediate fractions of $\sim 65\%$. For alloys with very small SFE there is another abrupt drop in the volume fraction of the $\langle 111\rangle$ fiber. However, one of the largest effects on the final texture is in fact the starting texture, which is often not measured and incorrectly assumed to be uniform. A nonrandom starting texture also tends to

Table 1
Textures commonly formed under axisymmetric deformation.

Crystal	Tension	Compression
f.c.c.	$\langle 111\rangle+\langle 100\rangle$	$\langle 110\rangle$
b.c.c.	$\langle 110\rangle$	$\langle 111\rangle+\langle 100\rangle$
h.c.p.	$\langle 10\bar{1}0\rangle$	$\langle 0001\rangle$

lead to a nonuniform distribution of crystallographic directions orthogonal to the deformation axis.

It should be noted that Table 1 shows that the texture resulting from tensile deformation of f.c.c. polycrystals is the same as that due to compression of b.c.c. polycrystals, and vice versa. This is a result of the transposition of the $\langle 110\rangle\{111\}$ slip systems of the f.c.c. metals to $\langle 111\rangle\{110\}$ in b.c.c. metals, although for the b.c.c. case slip on planes other than $\{110\}$ can also occur.

2. *Rolling Deformation Textures*

Rolling is probably the most important industrial deformation processing route as a great volume of metallic material is produced as sheet and plate, which may undergo further processing such as deep drawing to form beverage cans. The effects of rolling on textures have been studied extensively particularly with regard to aluminum alloys and steels. Typical rolled products may undergo several stages of both hot and cold rolling, which impart texture and an associated anisotropy in the physical properties of the sheet or plate.

Rolling is often approximated to be a plane strain deformation with the limited transverse flow taken to be zero. Crystal orientations with respect to the rolling geometry are generally quoted as $\{hkl\}$ $\langle uvw\rangle$, where $\{hkl\}$ are the Miller indices of the crystallographic plane parallel to the rolling plane, and $\langle uvw\rangle$ are the Miller indices of the crystallographic direction along the rolling direction (RD). The transverse direction (TD) is generally not quoted, but is readily obtained from the cross (or vector) product of the RD and normal direction (ND).

2.1 Face-centered Cubic Metals

The commonly observed rolling textures in f.c.c. metals and alloys are: copper $\{112\}\langle 11\bar{1}\rangle$, Taylor $\{4\ 4\ 11\}\langle 11\ 11\ \bar{8}\rangle$, S $\{123\}\langle 63\bar{4}\rangle$, brass $\{110\}\langle \bar{1}12\rangle$, and Goss $\{110\}\langle 001\rangle$. The positions of pole figure peaks corresponding to these textures are shown in Fig. 1 for both $\{111\}$ and $\{100\}$ poles. The majority of these texture components have two or four distinct (but symmetry related) variants (Fig. 1). For example $\{112\}\langle 11\bar{1}\rangle$ and $\{112\}\langle \bar{1}\bar{1}\bar{1}\rangle$ are the two variants of the copper texture related to each other by reflection in the ND–TD plane. Other orientations have been suggested in the literature over the years, and there is also some variability in the exact indices, and, therefore, orientation, used to describe a given component, in particular the S component.

This variability is, in part, due to the fact that a set of discrete specific orientations does not give a very good description of the texture. A better description is afforded by the use of "fibers" which are ranges of orientations that are linked by rotation about a common axis. Fibers, therefore, correspond to lines through orientation space. Two fibers correspond to the preferred orientations of rolled f.c.c. metals. These are the α fiber which runs between brass $\{110\}\langle \bar{1}12\rangle$ and Goss $\{110\}\langle 001\rangle$ orientations and the β fiber between copper $\{112\}\langle 11\bar{1}\rangle$ and brass $\{110\}\langle \bar{1}12\rangle$ which also passes through the S orientation.

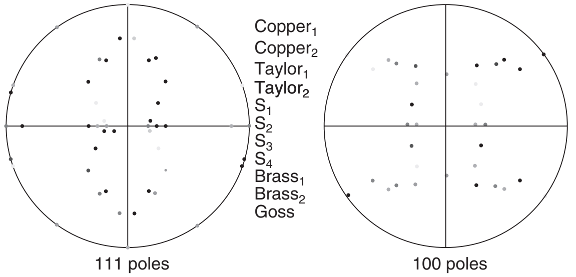


Figure 1
Common texture components for rolled f.c.c. metals and alloys (RD—horizontal, TD—vertical).

As the important orientations correspond to these fibers the presentation of experimental data can be simplified by showing the variation of texture intensity along the fibers rather than the full orientation distribution function (ODF). Often, however, the locations (in orientation space) of the preferred orientations do not lie exactly on the fibers just described, so that "skeleton lines" running close to the fibers through the peak texture intensities are used instead. Such use of skeleton lines increases the amount of information that must be presented as both the texture intensity variation along the skeleton line and its path through orientation space needs to be shown. This reduces the advantage of simplicity compared to showing the full ODF.

Rolling textures in f.c.c. metals and alloys can be broadly split into two categories: brass-like and copper-like (see *Annealing Textures*). Early observations showed that at room temperature copper develops a strong $\{112\}\langle 11\bar{1}\rangle$ texture (with some smaller amount of Goss orientation) sometimes referred to as a "pure metal" texture, while brass developed a distinct $\{110\}\langle \bar{1}12\rangle$ orientation also referred to as the "alloy" texture. It is now known that there is a steady change from copper-like to brass-like textures with increased alloying over a given finite composition width, and within the transition region both copper-like and brass-like texture components are observed to coexist. The brass-like texture is not found solely in alloys, however, as it has been observed in high-purity silver and other low-SFE metals. Conversely, high-SFE metals such as aluminum show a high tendency to form the copper-like rolling texture. Variation of the deformation temperature can also affect a transition in the rolling texture. For example, the texture of copper rolled at liquid nitrogen temperature is brass-like, while the brass-like texture of rolled silver becomes shifted toward that of copper if the rolling is undertaken at high temperature. Of course if the rolling is undertaken at sufficiently high temperature recrystallization and other annealing processes can occur resulting in textures that are a combination of a copper-like deformation component and a cube {100}⟨001⟩ recrystallization component (see *Annealing Textures*). In general conditions that promote planar slip, and inhibit cross-slip (low SFE, low temperature, shearable coherent precipitates) also promote the formation of brass-like texture.

There has been some debate in the literature on the possible link between the brass-like texture and the occurrence of deformation twinning. However, Kocks *et al.* (1998) point out that strong brass textures can form in systems such as Al–Li in which deformation twinning does not occur, and that the evidence for deformation twinning at the large strains associated with rolling is not clear.

2.2 Body-centered Cubic Metals

The overwhelming majority of work on rolling textures of b.c.c. polycrystals has been on steels because of their great technological importance. Rolling (and annealing) textures in steels have been reviewed by Ray *et al.* (1994). As with f.c.c. polycrystals the rolling textures are often analyzed in terms of fibers. The two most important of these are the α fiber occurring from {001}⟨1$\bar{1}$0⟩ to {111}⟨1$\bar{1}$0⟩ corresponding to a rotation about the RD (along ⟨1$\bar{1}$0⟩) and the γ fiber from {111}⟨1$\bar{1}$0⟩ to {111}⟨112⟩ corresponding to rotations about the ND (along ⟨111⟩). The paths through Eulerian orientation space (Bunge notation) of the α, γ, and other fibers are shown in Fig. 2. It is clear that a planar cut through the three-dimensional orientation space can be readily made so as to capture both α and γ fibers in a single section at $\phi_2 = 45°$. It has become common practice to just show this $\phi_2 = 45°$ section of the ODF when presenting textures of rolled b.c.c. polycrystals.

Cold rolling of many steels and other b.c.c. polycrystals leads to formation of α and γ fibers, with maximum intensities that move along the α fiber toward {112}⟨1$\bar{1}$0⟩ and along the γ fiber toward {111}⟨1$\bar{1}$0⟩ with increasing deformation. The details of the textures are somewhat dependent on alloy chemistry, e.g., the intensity of the strong {112}⟨1$\bar{1}$0⟩ texture is higher for niobium-stabilized IF steel (where it increases with niobium content) than for titanium-stabilized alloy (Ray *et al.* 1994). Another example given in Kocks *et al.* (1998) is the influence of silicon content on the rolling texture in a 0.05% C–0.20% Mn sheet steel. A significant change in texture is observed when the silicon content is raised to 3%. For steel with 2% Si content or less the α and γ fibers are present at similar intensities, but for 3% Si the γ fiber is somewhat suppressed and the α fiber increases intensity with a peak that moves from near {112}⟨1$\bar{1}$0⟩ to {001}⟨1$\bar{1}$0⟩. Carbon (solute) content and the size and morphology of precipitates has a relatively minor effect on the cold rolling textures of steels, though significant and technologically important effects are observed in subsequent annealing textures.

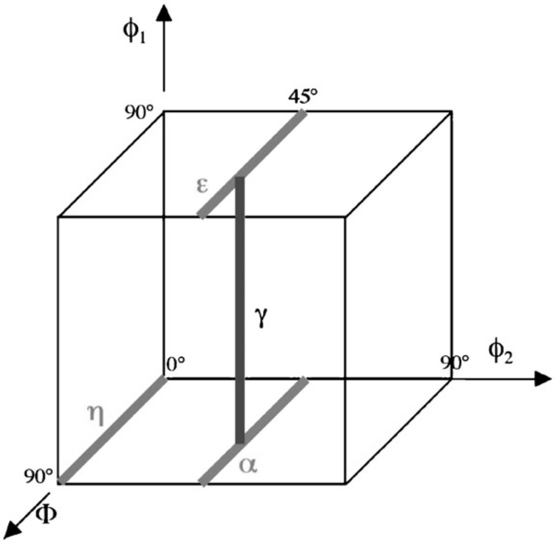


Figure 2
Positions of important fibers for rolled b.c.c. metals in Eulerian orientation space (Bunge notation).

Many of the effects seen in cold rolling of steels are significantly influenced by the texture and grain structures produced during prior hot rolling e.g., Ray *et al.* (1994) discuss how larger grain sizes developed during hot rolling lead to weaker partial α fiber on subsequent cold rolling. The hot rolling of steels is undertaken while in the austenite (γ) phase, but on cooling this transforms to the ferrite (α) phase. The transformation leads to an orientation relation between the prior γ and resulting α phases. This can be described by the well-known Kurdyumov–Sachs relationship which states that a $\{110\}_\alpha$ plane of the ferrite is aligned with a closed packed $\{111\}_\gamma$ plane of the prior austenite, while the $\langle 1\bar{1}1 \rangle_\alpha$ direction in the ferrite are parallel to the $\langle 1\bar{1}0 \rangle_\gamma$ direction in the austenite. This geometric relationship allows several ferrite variants to form from a given prior austenite grain. If the rolling temperature is sufficiently high then the austenite has a cube {100}⟨100⟩ recrystallization texture from which a weak rotated cube {001}⟨110⟩ texture results in the ferrite. Note that dominance of {001}⟨110⟩ over {110}⟨001⟩ and {110}⟨1$\bar{1}$0⟩ implies that some preferential selection of one of the Kurdyumov–Sachs variants occurs. When f.c.c. rolling textures are retained in the austenite (by lowering the temperature, or alloying with e.g., niobium or titanium) the transformation to ferrite results in hot rolling (termed hot banding) textures with significant intensity on the γ fiber.

2.3 Hexagonal Close Packed Metals

The properties of h.c.p. crystals are often quite markedly anisotropic, e.g., the elastic modulus for shearing on the prismatic planes of zinc (high c/a ratio) is 64% higher than for shearing on the basal planes. The plasticity of h.c.p. metals is also highly anisotropic and strong textures are frequently encountered.

The rolling textures of h.c.p. polycrystals are perhaps best categorized with respect to their c/a ratio (Fig. 3). The simplest textures form in materials that are close to the ideal c/a ratio of 1.633 such as magnesium. Here the basal planes aligns parallel to the rolling plane to form a [0001] fiber, in which there is no strong preferential alignment along the RD. For basal slip crystals having c/a ratios significantly greater than the ideal value (e.g., Zn) the [0001] poles are tilted away from the rolling plane normal by rotations of 15–25° about the TD, while the $[11\bar{2}0]$ poles align with the RD. For those metals with c/a ratios below the ideal value (prism/pyramidal slip), including the technologically important materials titanium and zirconium, the basal poles are again tilted away from the rolling plane ND. This time, however, the rotations tend to be slightly larger, 20–40°, and are about the RD to which the $[10\bar{1}0]$ poles align. The extent of the basal plane tilting which causes a splitting of the peak in the [0001] pole figure tends to increase with increasing cold rolling reduction. At higher temperatures the alignment of $\langle 10\bar{1}0 \rangle$ with the RD becomes less distinct in titanium, with an increasing tendency for $\langle 11\bar{2}0 \rangle$ to align with RD. Alloying can also affect the texture of h.c.p. polycrystals. For example, although solid solution substitutional alloying has little effect in titanium or zirconium, the introduction of interstitial atoms can lead to the formation of subsidiary peaks in [0001] pole figures with some basal planes being tilted by ~20° about the TD.

3. Modeling Deformation Texture Formation

3.1 Equilibrium and Compatibility Conditions

Predicting how individual crystals deform and rotate within the constraints imposed by neighboring crystals during plastic straining of polycrystalline aggregates is a difficult undertaking. It is clearly an oversimplification to assume that each crystal deforms independently of all others; that is to say as a single crystal under the same externally applied stress state (Sachs model). The different strains developed in the crystals would lead to voiding or interpenetration of crystals across their grain boundaries, which does not occur.

The requirement that the strains in neighboring grains are compatible with each other leads to the generation of additional local stresses (compatibility stresses) which modify the plastic flow within the grains. Since the deformation process can be taken to be quasistatic the local stresses must satisfy the equilibrium condition. This requirement that both strain compatibility and stress equilibrium must be satisfied across each boundary of every grain leads to a mechanics problem with no general analytic solution, so that some simplifying assumptions must be made.

3.2 Taylor Model

The Taylor model emphasizes the importance of the compatibility requirement, by imposing the assumption that the plastic strain is uniform throughout the polycrystal. Thus each grain, regardless of orientation,

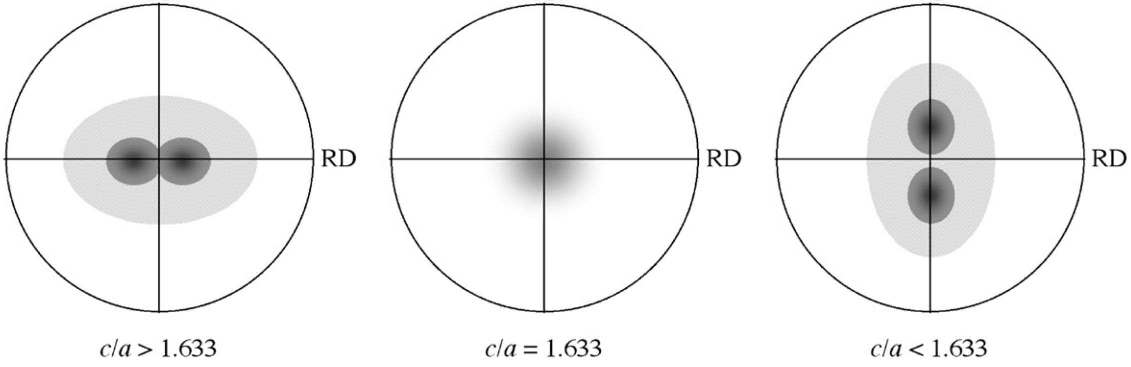


Figure 3
Effects of c/a ratio on pole figures showing basal plane orientations in cold-rolled h.c.p. polycrystals.

is forced to deform to the external macroscopic strain. Within each grain shearing on multiple slip systems accommodates the imposed strain and, in general, the action of at least five independent slip (or twinning) systems are required for an arbitrary strain. Generally many different combinations of different slip systems could achieve the required strain but Taylor (1938) proposed that the favored combinations were those that minimized the work done (shear stress × shear strain, summed over slip systems) against the plastic flow. This of course completely ignores the stress equilibrium requirements. This approach has been used to follow the reorientation of grains during successive strain increments so as to simulate the evolution of texture during deformation processing. A "relaxed constraint" modification of the theory has been developed in which the compatibility of strains is no longer imposed across boundaries that have become reduced in area due to the elongation of the grain.

The Taylor theory has been very successful, especially for highly symmetric cubic crystals, and reproduces the main features of deformation textures well, though it tends to predict somewhat sharper textures than are observed in practice.

3.3 Visco-plastic Models

An alternative approach to modeling polycrystal plasticity and texture effects is the visco-plastic self-consistent (VPSC) modeling scheme described fully in Kocks *et al.* (1998). In this treatment each grain is considered to be an ellipsoidal inclusion within a homogeneous effective medium. The properties of the homogenous effective medium are taken to be the average properties of the polycrystal, which are anisotropic. The misfit strain between the inclusion and the matrix leads to incompatibility stresses, which are obtained from Eshelby's (1957) analysis and are uniform throughout the inclusions (and independent of inclusion size, but not aspect ratio). The plastic deformation of the inclusion is obtained using nonlinear visco-plastic laws relating the shear-strain rate to the shear stress on each given slip system, which is linearized to obtain the strain increment for the current step along the strain path. Since the visco-plastic properties of the homogeneous effective medium are obtained from a suitable average of the properties of all the inclusions, it is not known *a priori*, but is calculated using a self-consistent iterative scheme.

Such treatments tend to emphasize stress equilibrium rather than strain compatibility. The textures predicted by VPSC simulations are in general agreement with observation, though the simulated ODFs tend to be less sharp, which is a result of underestimating the constraints imposed by neighboring grains. The VPSC treatment has some advantages over the Taylor model in that it can readily analyze lower symmetry materials (Wenk 1999) in which slip and twinning systems of considerably different shear strength exist, and it is easier to modify the calculation to incorporate interactions with specific neighbor grains.

See also: Annealing Textures

Bibliography

Bacon G E 1975 *Neutron Diffraction*. Clarendon Press, Oxford, UK

Bunge H-J 1987 Three-dimensional texture analysis. *Int. Mater. Rev.* **32**, 265–91

Dillamore I L, Roberts W T 1965 Preferred orientations in wrought and annealed metals. *Metall. Rev.* **10**, 271–380

Eshelby J D 1957 The determination of the elastic field of an ellipsoidal inclusion, and related problems. *Proc. R. Soc.* **241A**, 376–96

Kocks U F, Tomé C N, Wenk H-R 1998 *Texture and Anisotropy: Preferred Orientations in Polycrystals and Their Effect on Materials Properties*. Cambridge University Press, Cambridge, UK

Ray R K, Jonas J J, Hook R E 1994 Cold rolling and annealing textures in low carbon and extra low carbon steels. *Int. Mater. Rev.* **39**, 129–72

Taylor G I 1938 Plastic strain in metals. *J. Inst. Met.* **62**, 307–24

Wenk H R 1999 A voyage through the deformed Earth with the self-consistent model. *Model. Simul. Mater. Sci. Eng.* **7**, 699–722

Wilkinson A J, Hirsch P B 1997 Electron diffraction based techniques in scanning electron microscopy of bulk materials. *Micron* **28**, 279–308

A. J. Wilkinson
University of Oxford, UK

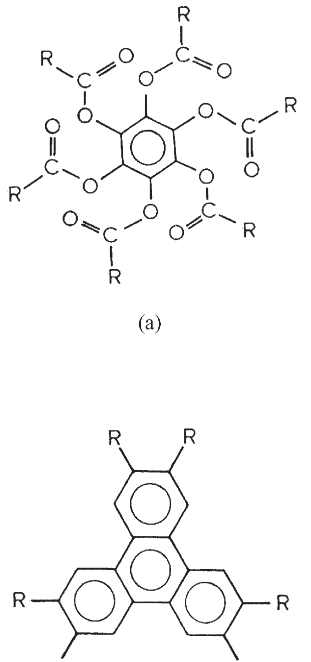
Discotic Liquid Crystals: Overview

Liquid crystals of disk-shaped molecules, now generally referred to as "discotic liquid crystals," were discovered in 1977 (Chandrasekhar *et al.* 1977). Since then, more than 1000 discotic compounds have been synthesized and a variety of mesophase structures identified. In their simplest form, these mesogens have flat or nearly flat cores surrounded usually by six or eight (sometimes four or seven) long-chain substituents. More complex discotic molecules are also known, (see, e.g., Chandrasekhar 1993). Available experimental evidence indicates that the presence of long side chains is crucial to the formation of these mesophases.

1. Description of the Liquid-crystalline Structures

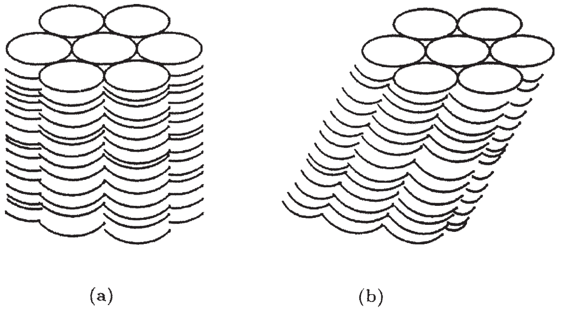
1.1 Columnar Liquid Crystals

The first examples of thermotropic mesomorphism in discotic systems were observed in hexaalkanoyloxy benzenes (Fig. 1(a)) and hexaalkoxy- and hexaalkanoyloxy triphenylenes (Fig. 1(b)) and it was established by x-ray studies (Chandrasekhar *et al.* 1977, Levelut 1979) that the mesophases of these compounds have a columnar structure. The basic structure of the columnar phase is illustrated in Fig. 2(a): the disks are stacked one on top of the other to form columns, the different columns constituting a two-dimensional lattice. Several variants of this structure have been identified—upright columns (Fig. 2(a)), tilted columns (Fig. 2(b)), hexagonal lattice, rectangular lattice, etc. In some cases, the columns are liquid-like i.e., the molecular centers are arranged aperiodically within each column, as depicted in Fig. 2(a) and 2(b), while in others they are arranged in a regular, ordered fashion or, more precisely, with a much larger range of positional correlation along the column axis. Other modifications are the hexagonal plastic columnar phase in which the disklike molecules undergo intracolumnar rotations about the columnar axis while retaining the three-dimensional positional order of their molecular centers (Glusen *et al.* 1996, 1997); and a highly ordered columnar phase which has a helicoidal stacking of the cores within each column (Fig. 3).



(b)

Figure 1
(a) Hexa-alkanoyloxy benzenes, $R = C_nH_{2n+1}$ (b) hexa-alkoxy- and hexa-alkanoyloxy triphenylenes, $R = C_nH_{2n+1}O$ and $C_nH_{2n+1}CO \cdot O$). These are the first examples of disk-shaped molecules showing thermotropic mesomorphism.



(a) (b)

Figure 2
Schematic representation of columnar structures of disk-shaped mesogens; (a) upright columns and (b) tilted columns.

The columnar phase is designated as "col" with appropriate subscripts to distinguish the different modifications. For example, the subscript "h" stands for a hexagonal lattice of columns, "r" for a rectangular lattice, "ob" for an oblique lattice, "d" for a disordered (liquidlike) stacking of the molecules within each column, "o" for an ordered stacking, and "t" for tilted columns. Thus, col_{hd} stands for columnar, hexagonal, disordered; col_{rd} for columnar, rectangular, disordered; col_{ho} for columnar, hexagonal, ordered; col_{t} for columnar, tilted; col_{ob} for columnar, oblique; col_{p} for columnar, plastic; col_{hel} for columnar, helical; etc. The three-dimensionally ordered col_{ho} are often referred to as "soft crystals" but optical textures and other studies reveal that they are different from solid crystals, and are considered as belonging to the family of liquid crystals (like the three-dimensionally ordered smectic B, smectic H, etc.).

Figure 4 presents a plan view of some of the different types of two-dimensional lattice structures that have been identified (Levelut 1983).

High-resolution x-ray studies have yielded important additional information about these structures. The measurements were made on well-oriented monodomain samples obtained by preparing freely suspended liquid crystal strands (typically 100–200 μm in diameter and 1.5 mm long) with the columnar axis parallel to the axis of the strand. The studies show that within each column, the flat molecular cores are approximately parallel to one another and orientationally well ordered, whereas the flexible hydrocarbon chains are in a highly disordered state. Similar conclusions can be drawn from deuterium NMR spectroscopy (see Chandrasekhar and Ranganath 1986).

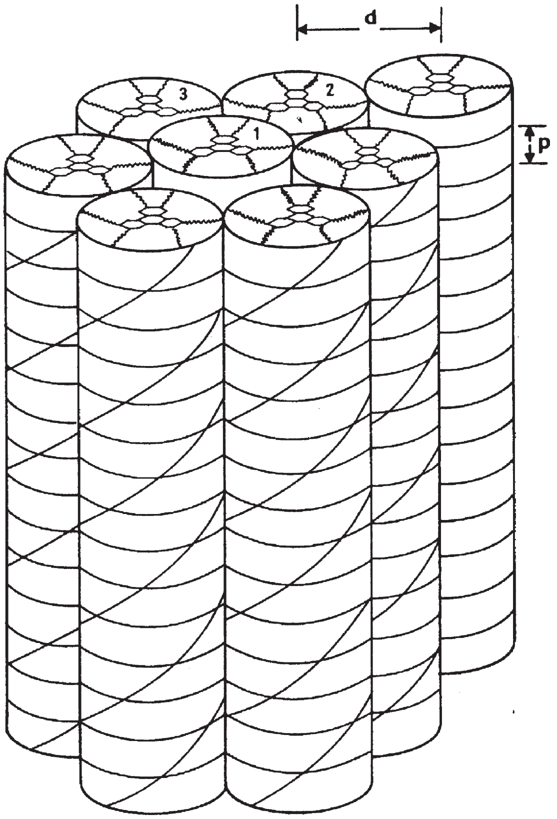


Figure 3
Structure of the col$_{\mathrm{hel}}$ phase of hexa hexyl thiotriphenylene, $R = SC_6H_{13}$ (HHTT): The diagram shows a three-column superlattice (not to scale) with column 1 displaced by p/2 relative to columns 2 and 3; neighboring columns have correlated helical phases (Fontes *et al.* 1988).

1.2 Columnar Phases of Nondiscotic Molecules

Columnar phases are also formed when the flat core is replaced by a conical or pyramidal shaped one (Fig. 5), and rarely when the central core is absent altogether, as in certain macrocyclic molecules (Fig. 6(a)). In the latter case, the columns are in the form of tubes (Fig. 6(b)).

Mesogens shaped like stick insects (Fig. 7(a)), called phasmids, form columnar mesophases, the structure of which is shown in Fig. 7(b).

Historically, an interesting case is that of diisobutyl silane diol (DIBSD, Fig. 8). The compound shows a mesophase which could not be classified as belonging to any of the then-known liquid crystal types. Twenty-five years later, Bunning *et al.* (1980) carried out miscibility studies of DIBSD with the newly discovered columnar phase of hexa-n-heptanoate of benzene, and proved conclusively that the mesophase is a columnar liquid crystal. It has been suggested that the DIBSD molecule forms a dimer or a polymer that favors the occurrence of a columnar phase. (For other interesting examples, including the formation of a cubic phase by discotic mesogens, see Chandrasekhar 1993.)

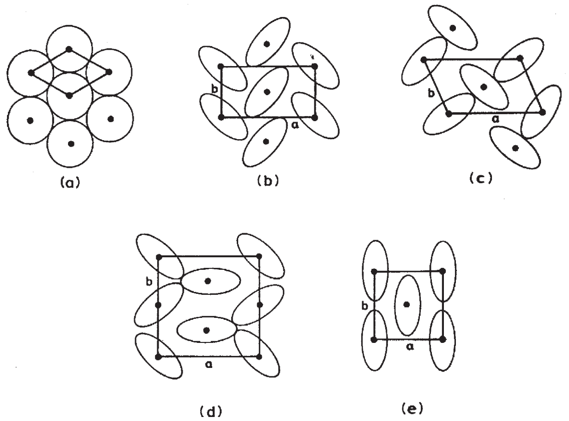


Figure 4
Plan views of the two-dimensional lattices of columnar phases; ellipses denote disks that are tilted with respect to the basal plane: (a) hexagonal ($P_62/m\ 2/m$); (b) rectangular ($P2_1/a$); (c) oblique (P_1); (d) rectangular (P_2/a); (e) rectangular, face-centered, tilted columns (C_2/m). (These are planar space groups when symmetry elements relating to translations along one of the axes, in this case the column axis, are absent (Levelut 1983)).

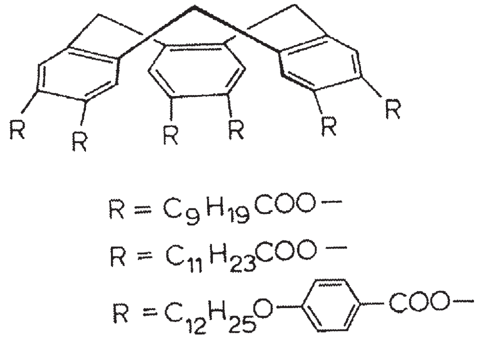
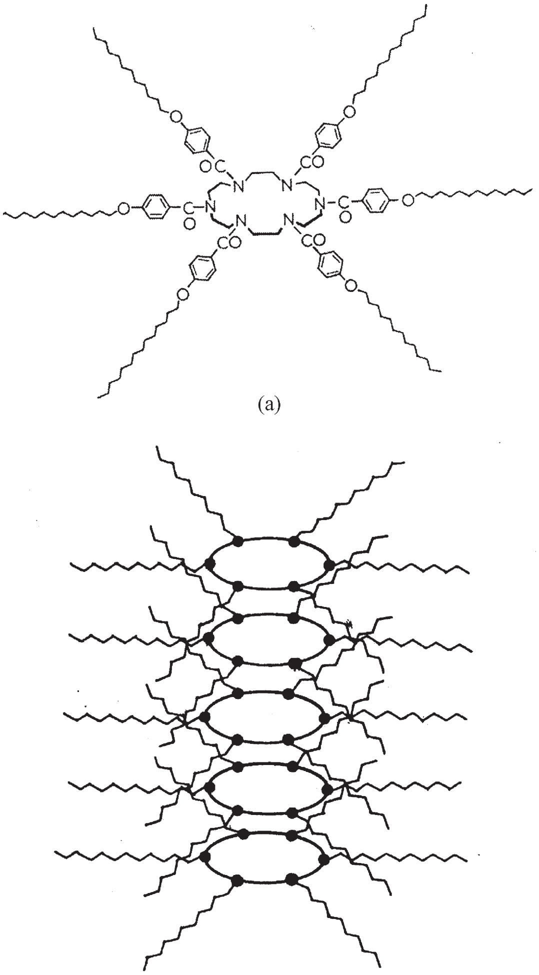


Figure 5
Cyclotricatechylene hexaesters, the cores of which are cone-shaped.

1.3 The Nematic Phase

The discotic nematic phase, designated by the symbol N_D, is a fluid phase showing schlieren textures similar to the classical nematic phase of rodlike molecules. Its structure is represented in Fig. 9(a): it is an orientationally ordered arrangement of disks with no long-range translational order. However, unlike the



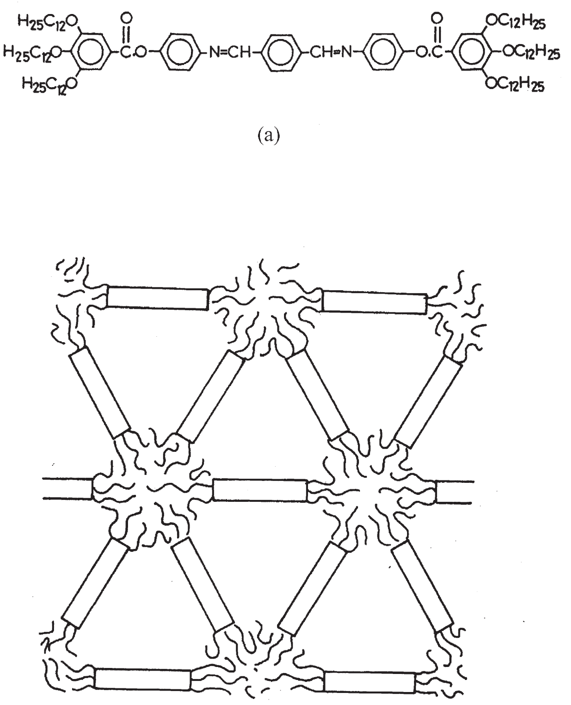
(a)

(b)

Figure 6
(a) hexa-(p-n-clodecyloxy benzoyl) derivative of macrocyclic polyamines, and (b) a representation of the columnar (tubular) mesophase.

usual nematic phase, N_D is optically negative, the director **n** now representing the preferred axis of orientation of the molecular short axis or the disk normal. Comparatively few compounds exhibit this phase; an example is shown in Fig. 9(b).

In some compounds (usually the lower members of the homologous series), the transition to the N_D phase takes place directly from the crystal, while in others (the higher members of the series), it takes place via a columnar phase. The trend is somewhat analogous to the behavior of the smectic A-nematic (A-N) transition in a homologous series of rodlike molecules. This trend can be explained by an extension of the well-known McMillan model of the A-N transition to discotic systems (see Chandrasekhar 1998).



(a)

(b)

Figure 7
(a) Hexa-n-alkoxy-terepthal-bis-[-4-benzoyloxy aniline] and (b) the structure of the hexagonal columnar mesophase formed by it. The molecule has been named "phasmids" (stick insects).

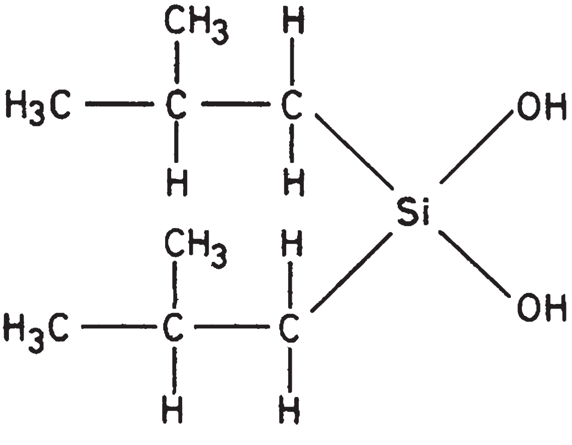
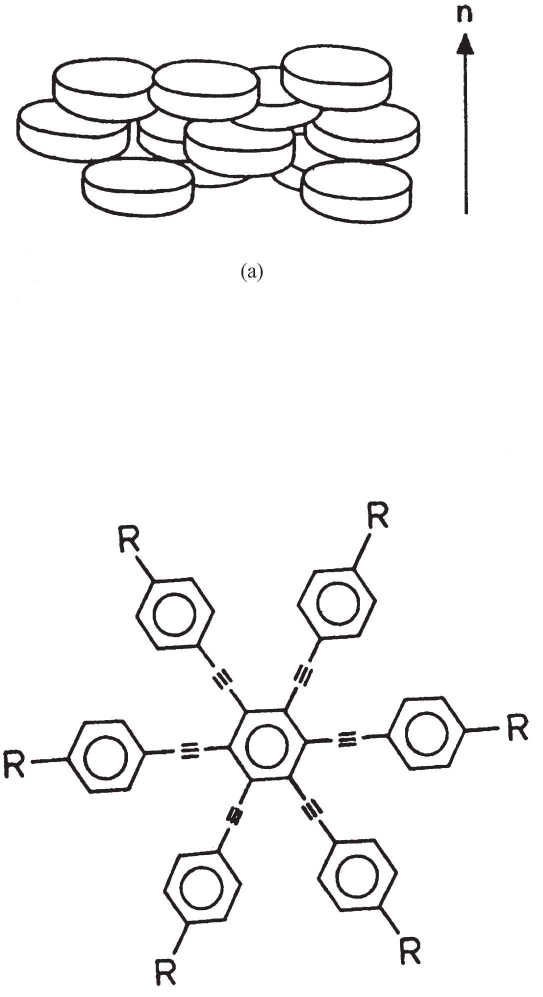


Figure 8
Di-isobutylsilane diol (DIBSD) which forms a columnar mesophase.



(a)

(b)

Figure 9
(a) Schematic representation of the molecular order in the discotic nematic (N_D) phase (b) hexakis-(4-octyl phenylethynyl)benzene, which forms the N_D phase (Kohne and Praefcke 1987).

The hexa-n-alkanoates of truxene (Fig. 10) show an unusual sequence of transitions. For the higher homologs, the phase sequence on cooling is Iso–$\mathrm{col_h}$–$\mathrm{col_r}$–N_D–Re-entrant $\mathrm{col_h}$–cryst, where Iso and cryst are the isotropic and crystalline phases respectively. It has been suggested that the truxene molecules are probably associated in pairs, and that these pairs break up at higher temperatures, which may explain this extraordinary behavior.

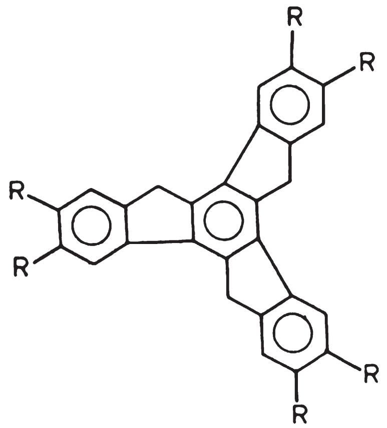


Figure 10
Hexa-n-alkanoates of truxene which exhibits a re-entrant sequence of transition.

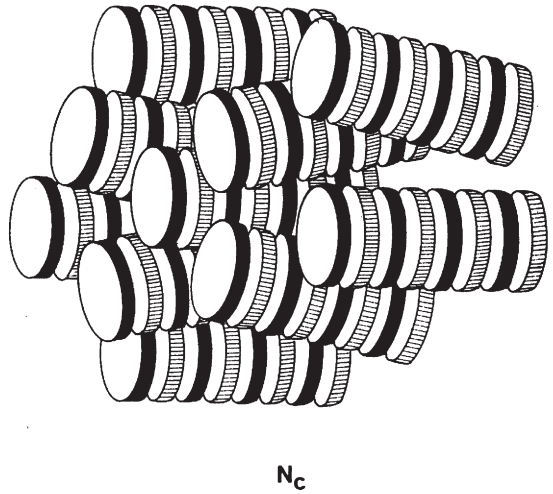


Figure 11
Schematic diagram of the structure of the columnar nematic phase, N_{col}, induced by charge transfer interaction between an electron donor (black) and an electron acceptor (hatched).

1.4 The Columnar Nematic Phase

The addition of electron acceptor molecules such as 2,4,7-trinitrofluorenone to electron-donating discotic molecules results in the formation of charge-transfer complexes which stabilize (or, in nonmesomorphic materials, induce) mesophases. A new type of mesophase observed in such systems is the columnar nematic (N_{col}) phase, the structure of which consists of short columns with the columnar axis approximately parallel to one another (Fig. 11).

1.5 The Chiral Nematic Phase

The chiral nematic, or cholesteric phase, of disk-shaped molecules occurs in mixtures of discotic nematics and chiral dopants, as well as in pure chiral discotic compounds. The phase, designated as N_D^*, has a structure as illustrated schematically in Fig. 12. The N_D^* phase shows reflection colors, very much like the usual cholesteric of rodlike molecules and exhibits typical cholesteric textures.

1.6 The Lamellar Phase

Giroud-Godquin and Billard (1981), who prepared and investigated the first discotic metallomesogen, bis(p-n-decylbenzoyl)methanato copper(II) (Fig. 13), suggested that its mesophase has a lamellar structure.

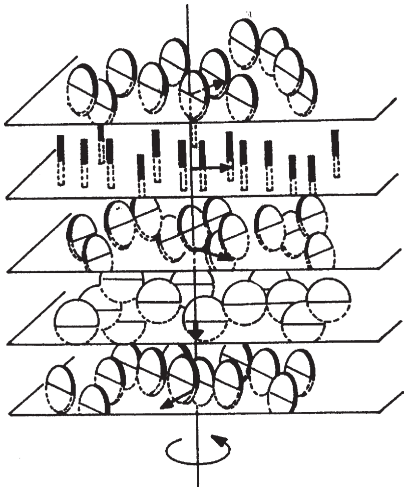


Figure 12
Schematic representation of the twisted nematic phase (N_D^*) of disklike molecules.

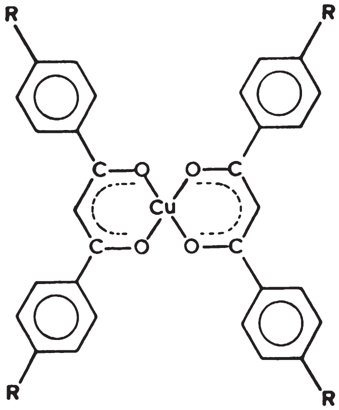


Figure 13
The molecular structure of bis (p-n-decylbenzoyl) methanato copper(II), the first example of a discotic metallomesogen.

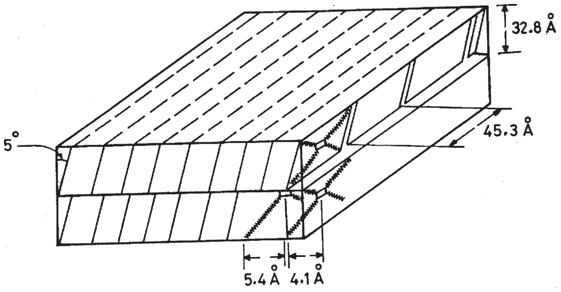


Figure 14
Schematic representation of the structure of the lamellar phase of the copper complex (Fig. 12) with $R = C_{12}H_{25}$ (proposed by Sakashita *et al.* 1988)

Similar conclusions have been drawn from x-ray studies (Ohta *et al.* 1986) and from NMR investigations (Ribeiro *et al.* 1988). Sakashita *et al.* (1988) have proposed a tilted smectic C-type of structure, as depicted in Fig. 14.

A variety of discotic metal complexes has been synthesized which exhibit columnar or lamellar phases (Giroud-Godquin and Maitlis 1991). X-ray determination of the crystal structures has been useful in elucidating the structures of the complexes and the mesophases formed by them (Usha *et al.* and references therein, 1993).

2. Physical Properties of Columnar Liquid Crystals

Some of the important properties of the columnar phases are summarized below:

2.1 One-dimensional Electrical Conductivity of the Columnar Phase

The separation between the aromatic cores within the column in these liquid crystals is often around 3.6 Å, so that there is considerable overlap of the Π-orbitals of adjacent molecular cores. One would therefore expect the electrical conductivity $\sigma_{\parallel}$ along the columnar axis to be high, but, in fact, the compounds are insulators. The reason for this is that their intrinsic charge concentration is low because of the large energy gap of 4 eV. However, the addition of dopants, electron acceptors, or donors, increases $\sigma_{\parallel}$ significantly—by as much as a factor of 10^7 (Balagurusamy *et al.* 1999). On the other hand, doping makes hardly any difference to the conductivity in the isotropic phase (Fig. 15). DC conductivity anisotropy $\sigma_{\parallel}/\sigma_{\perp}$ (where $\sigma_{\perp}$ is the electrical conductivity transverse to the columnar axis) as high as 10^{10} or more have been observed. Thus, the columns may well be described as "molecular wires."

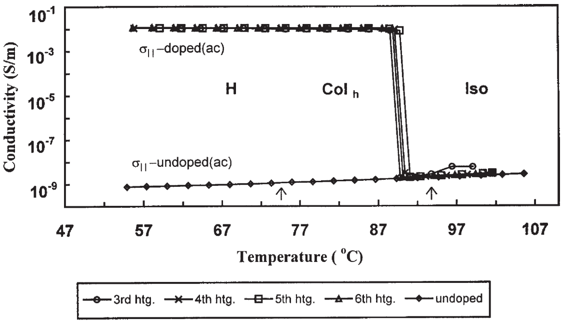


Figure 15
The measured AC electrical conductivity at 1 kHz parallel to the columns, $\sigma_{\|}$, in doped HHTT in the heating mode over several thermal cycles. The transition temperatures are indicated by the arrows.

The thermoelectric power of these "molecular wires" has been determined (Balagurusamy *et al.* 1999). The sign of the thermo-emf is an indication of the sign of the charge carriers in the doped compound, i.e., whether it is a p-type or n-type material.

2.2 Photoconductivity of the Columnar Phase

Similarly, very high values of mobility of photoinduced charge carriers (implying fast photoconduction) have been observed using a time-of-flight method along the columnar axes of ordered columnar phases, col_{hel}. Mobilities of the order of 0.1 $\text{cm}^2\ \text{V}^{-1}\text{s}^{-1}$ have been reported in the col_{hel} phase (Adams *et al.* 1994). With the exception of organic single crystals, these represent the highest values known to date of the mobility of photoinduced charge carriers in organic systems. These studies will have an impact on applications such as electrophotography and the development of new devices such as organic transistors and active layers in light emitting diodes.

2.3 Ferroelectricity

Columnar phases can exhibit ferroelectricity if the disk-shaped molecules are chiral and tilted with respect to the columnar axis (Bock and Helfrich 1995). From symmetry arguments, it follows that each column is spontaneously polarized perpendicular to the plane containing the columnar axis and the disk-normals. Macroscopic polarization is observed when the polarization of the different columns do not cancel each other. Switchable ferroelectric columnar liquid crystals have some advantages over ferroelectric smectics—greater resistance to mechanical shocks and material flow, nonappearance of chevron structures (which is a drawback in conventional (calamitic) ferroelectrics), a nearly temperature-independent tilt angle for the few cases that have been studied.

2.4 The Properties of Discotic Nematics

The symmetry of the discotic nematic N_D (Fig. 9(a)) is the same as that of the classical nematic of rodlike molecules. Thus, identical types of defects and textures are seen in both cases.

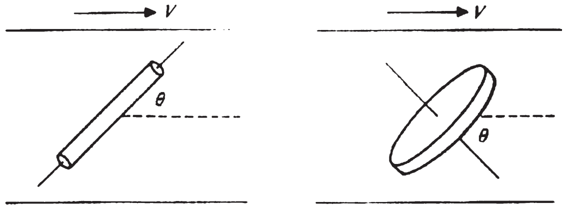
The N_D phase is optically, and, as a rule, diamagnetically, negative. However, its dielectric anisotropy $\Delta\varepsilon$ ($= \varepsilon_{\|} - \varepsilon_{\perp}$, where $\varepsilon_{\|}$ and $\varepsilon_{\perp}$ are the dielectric constants parallel and perpendicular to the director) may be positive or negative depending on the detailed molecular structure. For example, hexa-n-heptyloxybenzoate of triphenylene and hexa-n-dodecanoyloxy truxene are both dielectrically positive in the nematic phase (Mourey *et al.* 1982, Raghunathan *et al.* 1987). This is because, while the contribution of the electronic polarizability of these disk-shaped molecules $\Delta\varepsilon$ is negative, the permanent dipoles associated with the six ester linkage groups make a stronger positive contribution, so that $\Delta\varepsilon > 0$ for the two compounds. On the other hand, hexakis [(4 octyl phenyl)-ethynyl]benzene, which does not have any permanent dipoles, does in fact show negative $\Delta\varepsilon$, as expected (Heppke *et al.* 1988).

Very few quantitative measurements of the physical properties have been reported. The Frank constants for splay and bend have been determined using the Freedericksz method (Mourey *et al.* 1982, Raghunathan *et al.* 1987, Heppke *et al.* 1988, Warmerdam *et al.* 1987). Interestingly, the values are of the same order as for nematics of rodlike molecules. The twist constant k_{22} has not yet been measured. The fact that the diamagnetic anisotropy is negative makes it somewhat more difficult to measure these constants by the Freedericksz technique. However, by a suitable combination of electric and magnetic fields it is possible, in principle, to determine all three constants (Raghunathan *et al.* 1987).

The hydrodynamic equations of the classical nematic are applicable to the N_D phase as well. There are six viscosity coefficients (or Leslie coefficients), which reduce to five if one assumes Onsager's reciprocal relations. A direct estimate of an effective average value of the viscosity of N_D from a director relaxation measurement (Mourey *et al.* 1982; Warmerdam *et al.* 1987) indicates that its magnitude is much higher than the corresponding value for the usual nematic.

In ordinary nematics of rodlike molecules, the Leslie coefficients μ_2 and μ_3 are both negative and $|\mu_2| > |\mu_3|$. Under planar shear flow, the director assumes the equilibrium orientation θ_o given by

$$\tan^2\theta_o = \mu_3/\mu_2$$



(a) (b)

Figure 16
Flow alignment of the director in nematic liquid crystals. (a) For rod-shaped molecules, the alignment angle θ with respect to the flow direction lies between 0° and 45° (a), (b) while for disk-shaped molecules it lies between −90° and −45° (or equivalently between 90° and 135°).

where θ_o lies between 0° and 45° with respect to the flow direction (Fig. 16(a)). In practice, θ_o is usually a small angle. In certain nematics, it is found that $\mu_3 > 0$ at temperatures close to the nematic–smectic A transition point. Under these circumstances, there is no equilibrium value of θ_o, and in the absence of an orienting effect due to the walls or a strong external field, the director starts to tumble and the flow becomes unstable. Consequently, such materials have two distinct flow regimes which have been investigated in detail theoretically and experimentally (Chandrasekhar and Kini 1982).

The suggestion has been made that in the N_D phase, the disklike shape of the molecule may have a significant effect on μ_2 and μ_3 (Carlsson 1983, Volovik 1980). The stable orientation of the director under planar shear will now be as shown in Fig. 16(b). Thus, it can be argued that both μ_2 and μ_3 should be positive, and the flow alignment angle θ_o should lie between −45° and −90°. It then follows that when $\mu_2 < 0$, the director tumbles and the flow becomes unstable. As yet, however, no experimental studies have been carried out to verify any of these ideas.

3. Discotic Polymer Liquid Crystals

Another area of research which is attracting a great deal of attention since the discovery of discotics is a new class of liquid crystal polymers, viz. polymeric discotic liquid crystals.

Bibliography

Adams D, Schumacher P, Simmerer J, Haussling W, Paulus, Siemensmeyer K, Etzbach, K-H, Ringsdorf H, Haarer D 1994 Fast photoconduction in the highly ordered columnar phase of a discotic liquid crystal. *Nature* **371**, 141–3

Balagurusamy V S K, Krishna Prasad S, Chandrasekhar S, Sandeep Kumar, Manickam M, Yelamaggad C V 1999 Quasi-one dimensional electrical conductivity and thermoelectric power studies on a discotic liquid crystal. *Pramana* **53** (1), 3–11

Bock H, Helfrich W 1995 Two ferroelectric phases of a columnar dibenzopyrene. *Liq. Cryst.* **18**, 387–99

Bunning J D, Goodby J W, Gray G W, Lydon J E 1980 The classification of mesophase of di-i-butylsilanediol. In: Helfrich W, Heppke G (eds.) *Proceedings of Conference on Liquid Crystals of One- and Two-Dimensional Order, Garmisch-Partenkirchen*. Springer Verlag, pp. 397–402

Carlsson T 1983 Remarks on the flow alignment of disk-like nematics. *J. Phys.* **44**, 909–11

Chandrasekhar S 1993 Discotic liquid crystals: a brief review. *Liq. Cryst.* **14** (1), 3–14

Chandrasekhar S 1998 Discotic liquid crystals: their structure and physical properties. In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals*. VCH Verlagsgesellschaft GmbH Weinheim, Germany, 2B, pp. 749–80

Chandrasekhar S, Kini U D 1982 Instability in low molecular weight nematic and cholesteric liquid crystals. In: Ciferri A, Krigbaum W R, Meyer R B (eds.) *Polymer Liquid Crystals*. Academic, New York, Chapter 8. 202–45

Chandrasekhar S, Ranganath G S 1986 The structure and energetics of defects in liquid crystals. *Adv. Phys.* **35**, 507–96

Chandrasekhar S, Sadashiva B K, Suresh K A 1977 Liquid crystals of disc-like molecules. *Pramana* **9**, 471–80

Fontes E, Heiney P A, De Jeu W H 1988 Liquid-crystalline and helical order in a discotic mesophase. *Phys. Rev. Lett.* **61**, 1202–5

Giroud-Godquin A M, Billard J 1981 Un organometallique disquogene thermotrope. *Mol. Cryst. Liq. Cryst.* **66**, 147–50

Giroud-Godquin A M, Maitlis P 1991 Metallomesogens: metal complexes in organized fluid phases. *Angew. Chem. Int. Ed. Engl.* **30**, 375–402

Glusen B, Heitz W, Kettner A, Wendorff J H 1996 A plastic columnar discotic phase D_p. *Liq. Cryst.* **20**, 627–33

Glusen B, Kettner A, Wendorff J H 1997 A plastic columnar discotic phase. *Mol. Cryst. Liq. Cryst.* **303**, 115–20

Heppke G, Kitzerow H, Oestreicher G, Quentel S, Ranft A 1988 Electrooptical effect in a non-polar nematic discotic liquid crystal. *Mol. Cryst. Liq. Cryst. Lett.* **6**, 71–9

Kohne B, Praefcke K 1987 Liquid-crystalline compounds. 43. Hexaalkynylbenzene derivatives, the 1st hydrocarbon columnar-discotic or nematic-discotic liquid crystals. *Chimia* **41**, 196–8

Levelut A M 1979 Structure of a disc-like mesophase. *J. Phys. Lett.* **40**, L-81–L-84

Levelut A M 1983 Structure des phases mesomorphes formees de molecules discoides. *J. Chem. Phys.* **88**, 149–61

Mourey B, Perbet J N, Hareng M, Le Berre S 1982 Physical parameters of a fluid discotic mesophase. *Mol. Cryst. Liq. Cryst.* **84**, 193–9

Ohta K, Muroki H, Takagi A, Hatada K I, Ema H, Yamamoto I, Matsuzaki K 1986 Discotic liquid crystals of transition metal complexes, 3: the first-established discotic lamellar phase in bis[1,3-di(p-n-alkoxyphenyl)propane-1,3-dionato] copper(II). *Mol. Cryst. Liq. Cryst.* **140**, 131–52

Raghunathan V A, Madhusudana N V, Chandrasekhar S, Destrade C 1987 Bend and splay elastic constants of a discotic nematic. *Mol. Cryst. Liq. Cryst.* **148**, 77–83

Ribeiro A C, Martins A F, Giroud-Godquin A M 1988 Evidence of a smectic phase of disc-like molecules. *Mol. Cryst. Liq. Cryst. Lett.* **5**, 133–9

Sakashita H, Nishitani A, Sumiya Y, Terauchi H, Ohta K, Yamamoto I 1988 X-ray diffraction study on discotic lamellar phase in bis[1,3-di(p-n-alkoxyphenyl)propane-1,3-dinato]copper(II). *Mol. Cryst. Liq. Cryst.* **163**, 211–9

Usha K, Vijayan K, Chandrasekhar S 1995 Structural characteristics of some metallo-organic dimesogens. *Liq. Cryst.* **31**, 575–89

Volovik G E 1980 Relationship between molecule shape and hydrodynamics in a nematic substance. *JETP Lett.* **31**, 273–5

Warmerdam T, Frenkel D, Zijlstra J J R 1987 Measurements of the ratio of the Frank constants for splay and bend in nematic of disc-like molecules. *J. Phys.* **48**, 319–24

S. Chandrasekhar
Centre for Liquid Crystal Research, Bangalore, India

This Page Intentionally Left Blank

F

Ferrous Alloys: Overview

Steel is essentially an alloy of iron with (usually) a few tenths of 1 wt.% of carbon dissolved in interstitial sites in whichever iron crystal structure is present. The phase transformations and mechanical properties of steel can be substantially modified by the addition of widely varying proportions of alloying elements replacing atoms of iron in the iron crystal lattice. Steel is the most widely used metallic material in the world. Approximately 800 Mtyr^{-1} are produced (Paxton 1997). There are a number of reasons for these remarkable figures. Usually found as a mixture of oxides, but also with some sulfides, deposits of iron ore that can be economically mined are fairly well distributed around the world. Use of very large machinery has made the mining process economical when the ore is Fe-rich and can be mined from the surface rather than underground. Although extraction of iron from its ore is a difficult process, the technology involved is so well developed that the price of a kilogram of Fe alloyed with a fraction of a percent of C and $\sim$1% Mn is less than $0.50 (Paxton 1997). Highly alloyed steels, used for demanding special purposes, are priced an order of magnitude higher and hence are used only where low alloy steels are unable to meet the performance requirements. Steel strength can be varied from ca. 3500 MPa (500000 psi) down to ca. 35 MPa (5000 psi) through variations in chemistry, principally carbon concentration, and in heat treatment. Similar variations in ductility can be produced. With suitable alloying, Fe-base alloys can be made highly resistant to oxidation and to corrosion even at temperatures of several hundred degrees centrigrade

Upon heating from room temperature, "pure" Fe undergoes a phase transformation from b.c.c. (denoted as "α" or "ferrite") to f.c.c. ("γ" or "austenite") at temperatures above 912 °C. Above 1394 °C, Fe reverts to the b.c.c. structure, denoted as "δ" but again described as ferrite. At 1538 °C, δ Fe melts. As shown in Fig. 1, austenite can dissolve a large proportion of C (up to 2.1%) whereas ferrite can at most dissolve 0.021% C. Fe–C alloys containing less than the eutectoid carbon concentration (0.65%) are described as *hypoeutectoid*; at higher C contents, they are termed *hypereutectoid.* The pro-eutectoid ferrite reaction, occurring within the $\alpha+\gamma$ region as well as at lower temperatures, underlies much of the great utility of steel because ferrite is inherently very ductile except at low (subambient) temperatures but can be considerably strengthened by precipitation. A number of elements which dissolve substitutionally in both austenite and ferrite, such as manganese, nickel, chromium, and molybdenum can markedly affect the kinetics of this and other phase transformations in steel and may alter the mechanical properties associated with them.

1. Extractive Metallurgy and Mechanical Processing of Steel

Iron ore consists primarily of hematite (Fe_2O_3) and magnetite (Fe_3O_4) in proportions which vary appreciably among ore bodies, as well as variously hydrated hematites, siderite ($FeCO_3$), and pyrite (FeS_2). Among impurities present, again in widely varying proportions, the most important are lime (CaO), silica (SiO_2) and alumina (Al_2O_3). Hence extraction of Fe from its ore begins with reduction of the ferrous oxides by the addition of carbon or coke to the molten ore, usually in a blast furnace. Limestone is added to produce a predominantly lime–silica molten slag floating atop the steel bath.

The slag serves to protect the steel from re-oxidation and also to remove impurities whose solubility in the slag is greater than in the ferrous melt. The result of this process is known as "pig iron." This material contains 3–4.5% C, as well as about 1% Si and 1.5–2% Mn. In the next processing step, often undertaken in an electric furnace or a basic oxygen furnace, the carbon concentration is reduced by two orders of magnitude or more through the controlled passage of high-pressure oxygen through the melt. Substitutional alloying elements may be introduced through additions to a large refractory-lined ladle in which the molten steel is temporarily stored. Current practice is to solidify the steel through "continuous casting," in which an "endless" plate some cm. thick, or sheet as thin as 1 mm, is produced.

Recently, rolling of the product down to the desired thickness is being conducted on the bar/sheet as soon as it emerges from the continuous casting machine. Small heats of high alloy specialty steels, however, may be induction melted *in vacuo* or under a protective atmosphere and solidified by casting into ingots. They are subsequently forged, a process which can reduce porosity and other internal defects. An inevitable consequence of the very small solubilities in ferrite of oxygen (especially), sulfur, and phosphorous is the compounding of these (and other) elements into particles of oxides, sulfides and nitrides, in the melt and/or during austenite decomposition. These particles are termed "inclusions" when they appear in the final product. The properties of inclusions can exert significant effects upon mechanical properties, particularly at high-strength levels.

2. Phase Transformations in Steel

This section concerns alloys wholly transformable to austenite at elevated temperatures. The transformation

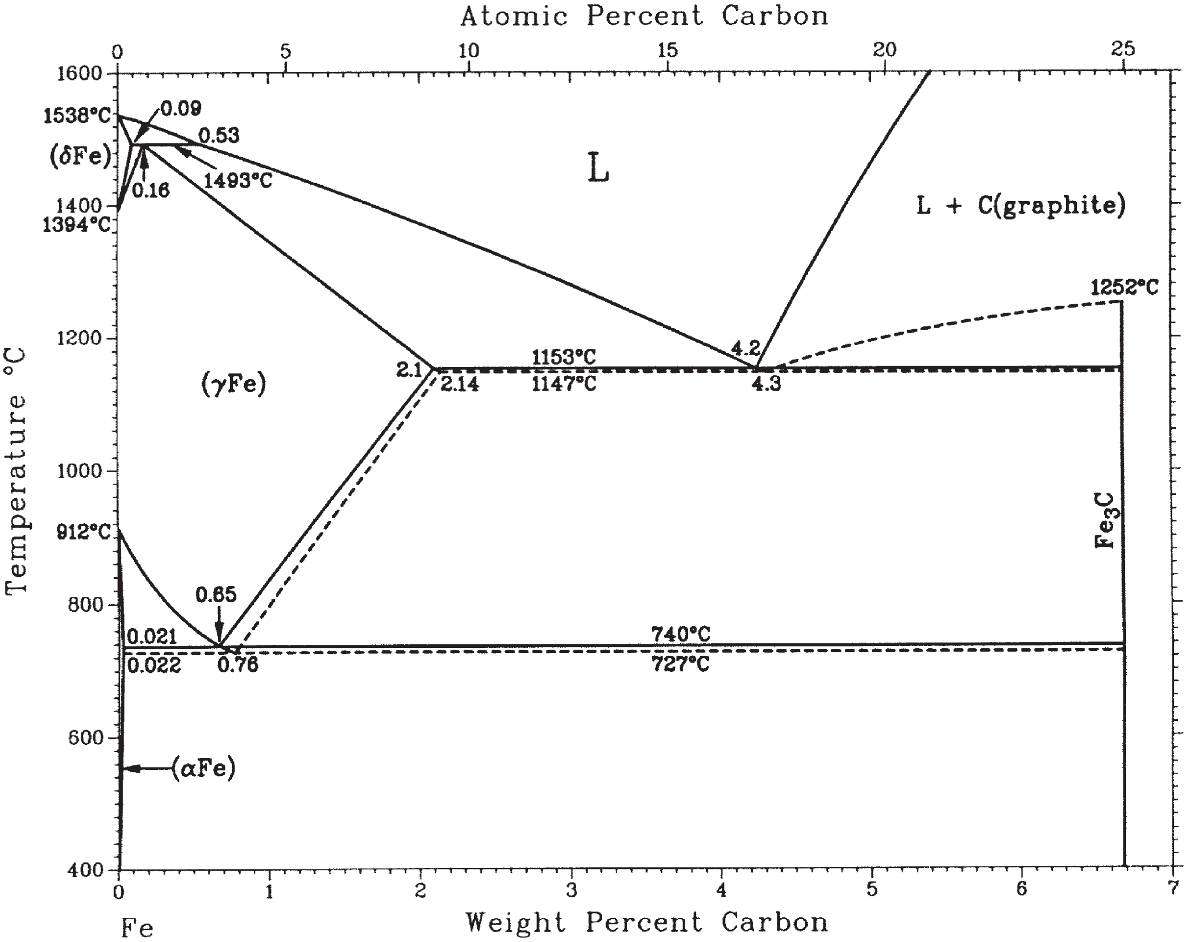


Figure 1
Fe-rich portions of the Fe–Fe_3C metastable equilibrium (solid lines) and Fe–C (graphite) equilibrium (dashed lines) phase diagrams (after Okamoto 1990).

processes considered involve decomposition of the austenite matrix at temperatures in the $\alpha+\gamma$ or the $\alpha+Fe_3C$ regions in Fig. 1 (and in the counterpart ranges when substitutional alloying elements are also present.) Alloys containing more than the maximum solubility of carbon in austenite (2.1% in Fe–C alloys; Fig. 1) are usually known as cast irons because they are too brittle to be shaped by plastic deformation. The considerations summarized here for steel are applicable in principle to cast iron, though there are major differences in detail, particularly because graphite particles can be present during both the liquid–solid and solid–solid phase transformations (as described in *Cast Irons*). Similarly, precipitation from ferrite and precipitation from austenite (see *Precipitation from Austenite*) are governed by the same principles but are also not discussed here.

Two basically different types of phase transformation take place in steels. At low temperatures, austenite transforms to martensite by a high-velocity lattice shearing process in which diffusion plays no role and in which the composition of the austenite is thus wholly inherited by the martensite. At higher temperatures, austenite decomposes through a variety of diffusional transformations. Although the latter processes are being increasingly found to resemble martensite in a crystallographic sense (Christian 1997), at the atomic level diffusional processes govern the very different mechanisms and kinetics operative.

Figure 2 is a time–temperature-transformation (TTT) diagram for a plain carbon steel. At temperatures above the martensite range, these plots are obtained by measuring the fraction of austenite transformed as a function of isothermal reaction time and

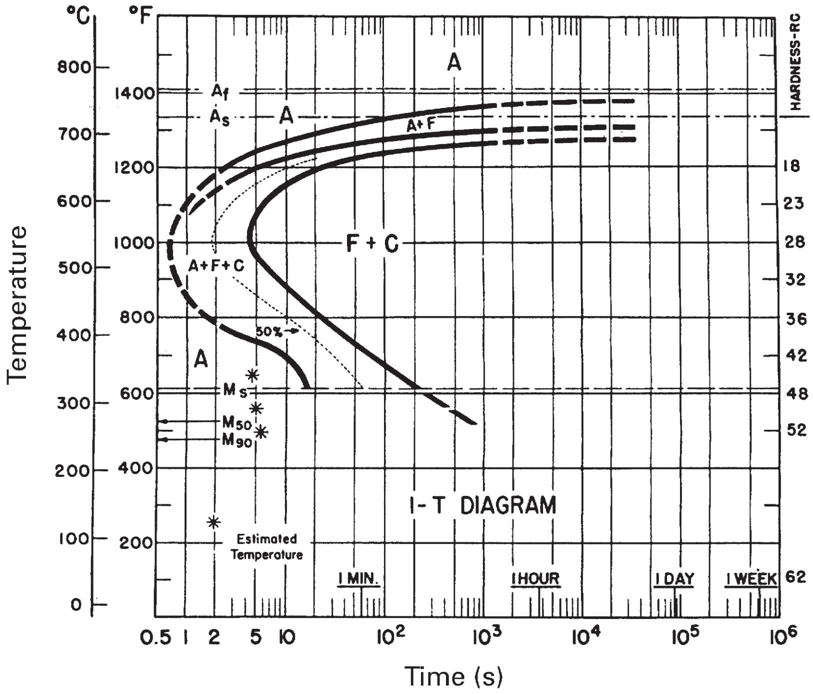


Figure 2
TTT diagram for initiation and completion of transformations in 0.50% C, 0.91% Mn steel. A = austenite, F = pro-eutectoid ferrite, C = carbide. A + F + C refers to pearlite above the "nose" of this diagram and to bainite at lower temperatures. M_s = martensite-start temperature (after ASM 1977).

determining the time required to begin a given transformation (e.g., at the 1% level) and to complete transformation of all austenite (e.g., 99%). These data are plotted as shown in Fig. 2. In eutectoid transformation processes, wherein both ferrite and Fe_3C (or an alloy carbide) form either co-operatively or competitively, the carbon diffusion distance tends to diminish rapidly with decreasing temperature, while the reverse situation obtains during the pro-eutectoid ferrite (and cementite) reactions.

Thus the TTT curves for the initiation of eutectoid reactions approach those for pro-eutectoid ferrite so closely at lower temperatures that the pro-eutectoid ferrite curve can no longer be delineated. The eutectoid reaction product at a temperature close to that of the eutectoid is usually pearlite (see *Pearlite*). This product consists of alternating lamellae of ferrite and Fe_3C which are forced to lengthen edgewise at the same rate because they share growth ledges passing through the edges of both phases (Hackney and Shiflet 1987a, 1987b). At lower temperatures, bainite appears (see *Bainite*). Although described by at least three mutually conflicting mechanisms, this product is taken here to form by the competitive growth of pro-eutectoid ferrite and carbide in nonlamellar form. The morphology of ferrite normally dominates that of the overall transformation product. At the temperatures where bainite occurs when the TTT diagram is similar to that of Fig. 2, the morphology of ferrite is often, though not always lath- or plate-like.

Since all of these reactions tend to nucleate predominantly at austenite grain boundaries (though certain inclusions can provide excellent intragranular nucleation sites, when the austenite grain size is reduced by use of low austenitizing temperatures and short austenitizing times, the presence of inclusions capable of restraining grain boundary migration or severe plastic deformation of the parent austenite which is not quickly recovered or recrystallized, the TTT curves are generally moved to shorter reaction times. At temperatures in the lower part of the bainite region, when sympathetic nucleation of ferrite at austenite:ferrite boundaries (Aaronson *et al.* 1995) predominates over grain boundary nucleation, this effect of austenite grain size can be largely nullified (see *Bainite*).

Usually, martensite forms only during cooling in steel, though it can do so isothermally in the absence of interstitials or in certain nonferrous alloy systems (see *Martensite*), the martensite reaction is represented on the TTT diagram by horizontal lines corresponding to difference percentages of martensite formed. In Fig. 2, M_s thus represents the "martensite-start" temperature, while M_{50} is the temperature by which 50% of austenite is transformed to martensite. Martensite formation can be expedited or hindered in the presence of elastic strains assisting or resisting the shear strain governing martensite formation. In steel, holding martensite at the transformation temperature, or especially, at a higher temperature, causes "tempering" by a variety of diffusional processes.

3. Mechanical Properties of Austenite Decomposition Products in Low Alloy Steels

Focussing first on steels with TTT diagrams similar to that in Fig. 2, some useful qualitative generalizations can be made about the mechanical behavior of austenite decomposition products. Pearlite formed at high temperatures, where the interlamellar spacing is coarse, is soft and also not particularly tough. However the rapid decrease in the interlamellar spacing as the reaction temperature is lowered permits the fine pearlite present near the nose of the TTT diagram to have a tensile strength of ca. 3500 MPa while retaining useful ductility. The presence of a few percent of grain boundary ferrite (allotriomorphs) can drastically decrease toughness when the matrix is transformed to martensite, as happens when a hypoeutectoid steel is quenched a little too slowly. A microstructure made up from large, roughly equiaxed ferrite crystals containing only widely spaced carbides is relatively weak.

However, as the ferrite grain size is decreased toward the 1 μm level, mechanical properties are markedly improved. Bainite formed at high temperatures, known as "upper bainite" (see *Bainite*), usually has poor toughness. This results from a low density of carbides and comparatively large regions or "colonies" in which all ferrite laths or plates are parallel. Alternatively, lower bainite usually has both a very high density of carbides and small regions within which sheaves of ferrite plates are parallel, thereby providing closely spaced barriers to dislocation movement. The toughness of lower bainite can rival or exceed that of martensite tempered to the same strength.

Unless the carbon concentration is $\leqslant$ca. 0.1%, martensite is very brittle. The hardness of martensite increases linearly with %C up to the eutectoid composition. With increasing tempering temperature and time, a variety of diffusional rearrangements of carbon occur within martensite. Precipitation of ε carbide ($Fe_{2.4}C$) and then of Fe_3C follow. In highly alloyed steels, the precipitation of alloy carbides takes place at still higher temperatures, where the diffusivity of alloying elements suffices to support this process. The latter type of precipitation can give rise to "secondary hardening" (see *Tool and Die Steels*), yielding strikingly high hardness and resistance to abrasive wear.

4. Mechanical Properties of Higher Alloy Steels

While low alloy steels comprise the bulk of all steel produced, more highly alloyed steels have been developed to enhance particular mechanical and related properties such as corrosion resistance, wear resistance, creep resistance, and toughness, or sometimes certain physical properties.

The most widely used alloy steels are the stainless steels. The most common stainless steels are the austenitic variety. Other stainless steels are the martensitic, the ferritic, the duplex (see *Stainless Steels: Duplex*), and the cast stainless steels (see *Stainless Steels: Cast*). Austenitic stainless steels have sufficient chromium to provide the desired corrosion and oxidation resistance and sufficient amounts of austenite stabilizing elements such as nickel/or manganese to ensure that the alloy remains austenitic at the expected service temperatures and that the austenite does not transform to martensite during loading in service. These alloys typically have low strengths and high formability.

The martensitic stainless steels can be austenitized and then quenched to form martensite. There are essentially three types of martensitic stainless steels. The first type contains carbon and can be strengthened by iron carbide precipitation at low tempering temperatures or by alloy carbide precipitation on tempering at higher temperatures. These steels have strength levels ranging from relatively modest (with carbon levels of 0.1 wt.%) to very high (at carbon levels of 0.8 wt.% or higher). The second type contains small amounts of carbon but are mainly strengthened by the formation of copper or intermetallic precipitates upon tempering.

The third type is strengthened by the precipitation of both alloy carbides and intermetallics. Ferritic stainless steels have the b.c.c. crystal structure and contain 11–30 wt.% chromium. These steels are less expensive than the austenitic steels and have excellent deep drawability and resistance to stress corrosion cracking. The duplex stainless steels contain alloying additions, primarily chromium and nickel, sufficient to provide corrosion resistance and 20–70 vol.% ferrite and 80–20 vol.% austenite. The duplex stainless steels have good resistance to oxidation, corrosion, and stress corrosion and have higher yield and ultimate tensile strengths than most austentic and ferritic steels. Cast stainless steels have been developed for two primary applications, corrosion resistance and high temperature properties including resistance to oxidation and to creep. The cast stainless steels developed for their high temperature properties are referred to as *heat resistant alloys* and have high carbon contents, unlike most wrought stainless steels.

Tool and die steels (see *Tool and Die Steels*) are used in cutting and forming operations. Tool steels tend to be low alloy steels but die steels are often medium carbon steels with sufficient amounts of strong carbide-forming elements such as chromium, molybdenum, tungsten, and vanadium to achieve pronounced secondary hardening. High-speed steels contain high concentrations of carbon and of strong carbide-forming elements. e.g., chromium, molybdenum, and tungsten. They can be used for high-speed cutting because they have excellent wear resistance at elevated temperatures. This resistance results from both secondary hardening and a large volume fraction of primary carbides.

The Hadfield steels were developed primarily for wear resistance. These materials are relatively high carbon Fe–Mn steels which are austenitic at room temperature but transform to a hard, strongly work-hardenable martensite when plastically deformed.

Two types of steel were developed to have high toughness. The first type is the martensitic steels having a low ductile-to-brittle transition temperature. Such steels, often called *cryogenic alloys*, have low strength but excellent toughness even at low temperatures. The second type are steels developed to have high toughness at very high strength levels at service temperatures close to room temperature but which do not have exceptionally low ductile-to-brittle transition temperatures. Such steels are the low-carbon maraging steels strengthened during tempering by precipitation of intermetallics or medium carbon, secondary hardening steels (such as AF1410) strengthened by alloy carbide precipitation.

Austenitic steels are the most resistant to plastic deformation at elevated temperatures (creep), but a number of creep resistant martensitic steels are also available. The first type of such steels are similar to the 0.15%C–2.25%Cr–1%Mo steels and the second are the 9–12% chromium steels containing 0.05–0.20 wt.% carbon. The 9–12 chromium steels typically have creep properties superior to those of the 2.25%Cr–1%Mo steels and can be used at higher temperatures.

Steels developed primarily for particular physical properties include the silicon steels, magnetic steels, and steels having low coefficients of thermal expansion.

Silicon steels are ferritic alloys whose magnetic properties are useful in motors and transformers. The silicon additions improve magnetic softness and increase electrical resistivity. Silicon also has the undesirable effects of decreasing the Curie temperature, reducing the saturation magnetization, and of embrittling the steel when the silicon $> \sim 2$ wt.%. Silicon steels are produced as highly textured grain-oriented alloys and also as alloys in which the grains are not oriented. Grain orientation is carried out to align the magnetically "easy" axis.

Iron-based alloys having near-zero or even negative expansion on heating in the vicinity of room temperature were discovered in ca. 1900 by the Swiss physicist Charles-Edouard Guillaume. The family of alloys possessing this characteristic is known as "Invar" for length INVARiant. The first Invar alloy had a composition of iron with 36 wt.% nickel. This alloy has a coefficient of thermal expansion varying from slightly negative to $< 10\%$ that of a normal metal from near absolute zero to near the Curie temperature. Invar exhibits a volume contraction as the alloy transforms from ferromagnetic to diamagnetic; this almost exactly counterbalances the concurrent thermal expansion. By varying the nickel content and alloying with additional elements Invar alloys have been developed which match the coefficients of thermal expansion of certain glasses and of silicon.

Bibliography

Aaronson H I, Reynolds W T Jr, Shiflet G J, Spanos G 1990 Bainite viewed three different ways. *Metall. Trans. A* **21**, 1343–80

Aaronson H I, Spanos G, Masamura R A, Vardiman R G, Moon D W, Menon E S K, Hall M G 1995 Sympathetic nucleation: an overview. *Mater. Sci. Eng. B* **32**, 107–23

ASM 1977 1977 *Atlas of Isothermal Transformation and Cooling Transformation Diagrams* ASM International, Materials Park, OH, p. 15

Christian J W 1997 Lattice correspondence, atomic site correspondence and shape change in diffusional-displacive phase transformations. *Prog. Mater. Sci.* **42**, 101–8

Hackney S A, Shiflet G J 1987a The pearlite:austenite growth interface. *Acta Metall.* **35**, 1007–17

Hackney S A, Shiflet G J 1987b Pearlite growth mechanism. *Acta Metall.* **35**, 1019–28

Okamoto H 1990 Fe–C phase diagram. In: Massalski T B (ed.), (2nd edn.) *Binary Alloy Phase Diagrams*. ASM International, Materials Park, OH, p. 843

Paxton H W 1997 Steel. In: *Encyclopedia of Chemical Technology, 4th edn*. Wiley, New York, 22 pp. 765–82

W. M. Garrison Jr
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

H. I. Aaronson
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA and Monash University, Clayton, Victoria, Australia

Fluorine-containing Polymers

Fluorine-containing polymers are organic polymers in which some or all of the hydrogen atoms bonded to carbon are replaced by fluorine. Because fluorine forms the strongest single bond to carbon and is the most electronegative element (Smart 1994), this substitution has profound effects on polymer properties. Highly fluorinated polymers have outstanding resistance to chemical attack, good thermal stability, and excellent weatherability, plus low dielectric constant, surface energy, moisture absorption, and flammability.

Although fluorine has a substantially larger van der Waals radius (1.47 Å) than hydrogen (1.20 Å), it is small enough to replace some or all hydrogen atoms without excessive crowding. Thus, the potential number of fluorine-containing polymers is vast and many structures have been reported. For example, fluorinated groups have been incorporated into polyesters, polyamides, and polyimides, often with the intent of increasing thermal stability or lowering the dielectric constant and surface energy (Cassidy and Fitch 1997, Feiring 1994), but these materials are not yet important commercial products. The direct fluorination of polymers, although not unknown (Anand *et al.* 1994), is rarely used; virtually all commercially important fluorinated polymers are produced by free-radical addition polymerizations of a few fluorinated monomers (Fig. 1). They include homopolymers of tetrafluoroethylene (TFE), 1,1-difluoroethylene (vinylidene fluoride, VDF), fluoroethylene (vinyl fluoride, VF), and chlorotrifluoroethylene (CTFE) and their copolymers with hexafluoropropylene (HFP), perfluoromethylvinyl ether (PMVE), and perfluoropropylvinyl ether (PPVE). Copolymers with nonfluorinated olefins, such as ethylene, propylene, and alkyl vinyl ethers, are also produced.

The particular combination and ratio of monomers used largely determine polymer properties, which can range from high melting, highly crystalline thermoplastics to fully amorphous elastomers (Yamabe *et al.* 1997). Ionomers are also produced by the incorporation of specific functionalized comonomers.

$CF_2{=}CF_2$	$CF_3CF{=}CF_2$	$CH_2{=}CF_2$
TFE	HFP	VDF
$CH_2{=}CHF$	$CClF{=}CF_2$	$CF_2{=}CFOCF_3$
VF	CTFE	PMVE
$CF_2{=}CFOC_3F_7$		
PPVE		

Figure 1
Monomers used in the major commercial fluoropolymers.

1. Thermoplastics

Selected properties of the key fluoroplastics are summarized in Table 1 (Feiring 1994).

1.1 Polytetrafluoroethylene

The longest known and most widely used fluoropolymer is polytetrafluoroethylene (PTFE). TFE polymerizes readily and can undergo a highly exothermic disproportionation, producing carbon and CF_4. Deflagrations can result, so safety is a key issue for producers. Typically, the monomer is manufactured and used at the same site, stored in minimum quantities, and inhibited with terpenes and inert diluents.

PTFE is produced by batch homopolymerizations of TFE in water at relatively modest temperatures (40–100 °C) and pressures (1–6 MPa), using a water-soluble free radical initiator such as ammonium persulfate. One process employs little or no surfactant and vigorous agitation to produce relatively large polymer particles, which are filtered and dried to give a granular material. A second process uses small amounts of a fluorinated surfactant, such as ammonium perfluorooctanoate, and gentle agitation to produce approximately 0.2 μm polymer particles in an aqueous dispersion. The dispersion may be used directly to produce coatings or coagulated to produce a product called fine powder.

The symmetry of TFE and the few imperfections, such as chain branches, formed during polymerization result in a virgin polymer with high crystallinity (>95%) and melting point (342 °C). PTFE objects would be brittle if this level of crystallinity was maintained in the final product. In practice, crystallinity is controlled by polymerizing to extremely high molecular weight, estimated at upwards of 5×10^7. The melted polymer can only crystallize to a limited extent on cooling, to give a material with a reproducible

Table 1
Properties of commercial fluorinated plastics.

Property	PTFE	FEP	PFA	ETFE	PVDF	PVF	CTFE	ECTFE
Melting point (°C)	327	265	305	270	180	190	210	240
Upper use temperature (°C)	260	200	260	150	150	110	180	150
Specific gravity	2.18	2.15	2.15	1.70	1.76	1.38	2.10	1.70
Dielectric constant	2.10	2.02	2.06	2.6	8–9	6–8	2.4	2.6
Dissipation factor	0.003	0.008	0.008	0.005	0.018	0.02	0.01	0.0024
Critical surface tension ($mN\,m^{-1}$)	18.5	17.8	17.8	22.1	25	28	31	

melting point of 327 °C. A consequence of this high molecular weight is an extremely high melt viscosity, so PTFE cannot be processed by typical thermoplastic techniques. The polymer is also insoluble, except to a very limited extent in a few perfluorocarbon solvents above its melting point, so solution processing is not a practical option. Thus, a number of specialized techniques had to be invented to convert the virgin polymer into useful forms (Blanchet 1997). For example, the granular polymer can be compressed and sintered to fuse the particles into dense objects, which can be machined into shapes or skived into sheet or tape. The fine powder can be mixed with a hydrocarbon lubricant to form a paste, which is extruded to form tubes, rods, or wire coatings. The extrudate is heated to drive off the lubricant and sinter the polymer particles.

PTFE can be used continuously up to 260 °C and is stable for shorter periods above its melting point. It will not burn in air, has a low-energy "non-stick" surface, and is resistant to attack by all chemicals, except for alkali metals, at normal use temperatures. It has an exceptionally low dielectric constant and dissipation factor and a high dielectric strength. Its tensile strength and modulus are modest, but it has good toughness and flexibility and high elongation for a crystalline thermoplastic. It does have a tendency to creep under ambient conditions, which must be considered when designing parts from PTFE.

The major application for PTFE is as electrical wiring insulation. Other important uses include pipe lining, valves, gaskets, and tubing for the chemical process industry and dispersion-coated fabrics for architectural applications. A specially stretched PTFE film with small pores, which transmit water vapor but are impermeable to liquid water, is used to produce weatherproof garments. Coated cookware is another major consumer application. Some PTFE is converted into fibers which are used as pump packing. PTFE "micropowders" have a lower molecular weight and are produced by deliberate synthesis or by radiation degradation of the standard resin. They are used as lubricity additives for oils and greases.

1.2 Tetrafluoroethylene Copolymers: FEP and PFA

PTFE is the least costly fluoropolymer, but its lack of melt processability limits the range of shapes which can be produced and adds to the complexity of parts manufacture. The need for materials which combine PTFE properties with melt processability is addressed by two commercial copolymers. Fluorinated ethylene propylene (FEP) and perfluoroalkoxy resin (PFA) are copolymers of TFE with about 10–12 wt.% hexafluoropropylene (HFP) or 2–4 wt.% perfluoropropylvinyl ether (PPVE), respectively. The degree of crystallinity is limited by the side chain "defects," so the copolymers can be produced with molecular weights of a few hundred thousand, more typical for thermoplastics.

FEP is generally produced by an aqueous dispersion process similar to that used for PTFE dispersions, although reaction temperatures (90–120 °C) and pressures (4–5 MPa) tend to be higher due to the lower reactivity of HFP. A lower temperature process uses a mixture of water and liquid HFP monomer as the polymerization medium. PFA is produced by aqueous, nonaqueous, and hybrid processes. The aqueous process is similar to that for PTFE dispersions. Nonaqueous processes are suspension polymerizations in an organic medium. They generally run at a lower temperature (30–60 °C) and use a perfluoroacylperoxide initiator. Hybrid processes employ a mixture of water and an organic solvent with an organic initiator. The organic solvent is traditionally a chlorofluorocarbon (CFC), such as 1,1,2-trichlorotrifluoroethane, but these solvents are being replaced by chlorine-free fluorocarbon solvents, because of depletion of the earth's ozone layer by the CFCs. Supercritical carbon dioxide is reported to be a suitable medium for conducting TFE copolymerizations (Romack *et al.* 1995).

FEP and PFA have chemical, electrical, and surface properties similar to those of PTFE. PFA has a melting point of about 305 °C, and its upper use temperature is the same as PTFE. FEP has a somewhat lower melting point (265 °C), and its upper use temperature is around 200 °C. Thus, PFA is often regarded as the premium product, but it is more expensive due to its PPVE comonomer. Both copolymers tend to have higher modulus and tensile strength than PTFE. They can be processed on conventional thermoplastic melt processing equipment, although the relatively high temperatures require corrosion-resistant parts.

The major use for FEP is in wiring insulation, especially for the telephone and computer cables running through the plenum spaces in commercial buildings. The dramatic growth in this application has made FEP the second largest fluoroplastic by volume. Other applications include release films and parts for the chemical process industry. PFA is used for constructing the wafer carriers, pumps, pipes, and filters that handle the ultrapure or corrosive materials employed by the semiconductor industry. Smaller amounts are used for similar applications in other industries and for electrical wiring insulation.

1.3 Other TFE Copolymers

TFE, an electron-poor monomer, tends to copolymerize readily with relatively electron-rich monomers, including ethylene, propylene, isobutylene, vinyl acetate, and alkyl vinyl ethers, to give materials with a wide variety of properties.

Copolymerization of TFE with ethylene provides a semicrystalline resin, ETFE. The two monomers tend to alternate, so the polymer repeating unit is largely–$CH_2CH_2CF_2CF_2$–. Commercial products also incorporate a few percent of a termonomer, such as perfluorobutyl(ethylene), hexafluoropropylene, or perfluoropropylvinyl ether, for improved mechanical properties (Kerbow 1997).

The manufacturing processes for ETFE are similar to those used for PFA and include nonaqueous, aqueous, and hybrid systems. The polymer is melt-processable using conventional equipment. It is not considered solution-processable, because it has only limited solubility in a few solvents, such as diesters, at temperatures close to its melting point.

Although incorporation of ethylene compromises some properties such as upper use temperature, dielectric constant (Table 1) and chemical resistance to strong oxidizing agents, ETFE has a lower density and significantly better toughness and abrasion resistance than the perfluoropolymers. ETFE can also be cross-linked by electron beam irradiation, or by heating with peroxides in the presence of cross-linking agents, to help maintain its mechanical properties at elevated temperatures.

A major application of ETFE is for wire and cable insulation, especially for use in environments such as aircraft, where superior mechanical properties are important. Other applications are in the chemical process industry to make or line items which must withstand aggressive chemicals.

Copolymers of TFE or chlorotrifluoroethylene with alkylvinyl ethers also tend to have alternating structures. These materials are amorphous and are soluble in moderately polar organic solvents, such as acetone or THF. Solution-based coatings for high performance automobiles and architectural applications have been developed from this family of fluoropolymers. A hydroxyalkyl vinyl ether termonomer is usually incorporated for curing (cross-linking) on application. Presence of the fluoromonomer in these compositions improves oxidative resistance, so the finishes remain glossy and attractive for a long time (Yamabe 1994).

1.4 Non-TFE Thermoplastics

Poly(vinylidene fluoride) (PVDF) is the most important non-TFE-based fluoroplastic and, in many respects, the most unusual fluoropolymer. Although its backbone structure, consisting mostly of–CH_2CF_2 CH_2CF_2– units, is isomeric with ETFE, its properties are quite different (Table 1). PVDF has a melting temperature nearly 100 °C lower and a much higher dielectric constant. PVDF dissolves at room temperature in polar organic solvents, and forms homogeneous blends with certain carbonyl-containing polymers. It has good chemical and oxidative resistance, but is far more sensitive to organic and inorganic bases than ETFE. The –CH_2 groups between two –CF_2 groups are relatively acidic so even mild bases, such as organic amines, can cause dehydrofluorination.

PVDF forms at least four different crystalline phases. The two most important are α, formed under normal crystallization conditions, and β, formed by crystallization under pressure or by mechanical deformation of polymer films. The former is the thermodynamically stable structure and consists of chains with a *trans-gauche* conformation. In the latter, the chains have a planar, zig-zag structure with fluorines on one side and hydrogens on the other. This polar structure is responsible for the piezoelectric and pyroelectric properties of PVDF.

PVDF is produced by aqueous and nonaqueous dispersion polymerization processes similar to other fluoropolymers. It is melt- and solution-processable. Major products include the homopolymer and copolymers with 5–12% of HFP for greater flexibility. Polymers containing TFE or both TFE and HFP are also produced, the latter having a relatively low processing temperature among the fluoroplastics. Copolymers of VDF with TFE are also of interest for improved piezoelectric properties.

A major application for PVDF is as an architectural coating (Iezzi 1997). Dispersions of PVDF are applied to metals such as steel or aluminum to provide a decorative and weather-resistant coating for commercial and residential buildings. It is used for electrical insulation, despite its higher dielectric constant, and in the chemical process industry. PVDF also has several unique small-volume applications in piezoelectric films for switches, sensors, and loudspeakers, ultrafiltration membranes for use in biotechnology, and as the electrode separator film in advanced batteries.

The least fluorinated among the major fluoroplastics is poly(vinyl fluoride) (PVF). It is produced by aqueous dispersion polymerization, but requires higher temperatures and pressures than other fluoropolymers. It has a somewhat irregular structure, due to chain branches and some head-to-head monomer incorporation, in addition to the expected head-to-tail enchainment. PVF has marginal stability above its melting point. It is processed by mixing with a high-boiling, polar organic solvent, such as tetramethylenesulfone, to form an organosol with the appearance of a fluid, single phase. This mixture is extruded into films and the solvent evaporated. The major application for PVF is as decorative and protective films for indoor and outdoor surfaces, such as the interior panels in commercial aircraft. It is also used as a release film for molding of thermoplastics.

Although it was first developed at about the same time as PTFE, the homopolymer of chlorotrifluoroethylene (PCTFE) is of minor commercial interest. PCTFE is melt-processable and much less permeable to gases than PTFE, but has lower melting point, a lower upper use temperature, poorer chemical resistance, and higher surface energy. The monomer is also significantly more expensive than TFE. A copolymer of CTFE with a few percent of VDF has improved processability.

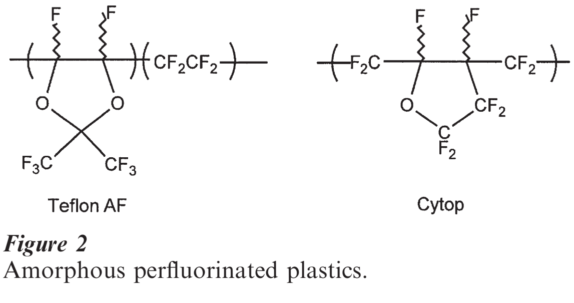
Of greater commercial significance is the copolymer of CTFE with ethylene. ECTFE has a higher melting point than either of its constituent homopolymers. The polymer has many similarities to ETFE in preparation, properties, and applications. Like ETFE, the two monomers are nearly alternating and the polymer has better mechanical properties than the perfluoropolymers. ETFE tends to have slightly better thermal stability and chemical resistance, whereas ECTFE leads in mechanical and barrier properties.

1.5 Amorphous Thermoplastics

Most fluoroplastics are semicrystalline materials with poor solubility. There are two amorphous perfluoroplastics which combine the outstanding thermal, chemical, and surface properties of PTFE with unique optical and solubility characteristics. Both polymers are produced by free-radical addition polymerization and contain rings in their backbone structures, but the approaches to ring formation are different.

Teflon AF (Fig. 2) refers to a family of random copolymers of TFE and a cyclic monomer, perfluoro-(2,2-dimethyldioxole) (PDD) (Resnick and Buck 1997). Glass transition temperatures (T_g) range from 335 °C for the PDD homopolymer to about 80 °C for a copolymer containing 20% PDD. The PPD homopolymer is difficult to process, so commercial products are copolymers containing 65% and 80% PPD with T_g values of about 160 °C and 240 °C, respectively. The second amorphous perfluoropolymer, Cytop (Fig. 2) is obtained by cyclopolymerization of perfluoro(butenyl vinyl ether) (Sugiyama 1997). This product has a T_g of 108 °C.

Teflon AF and Cytop have good solubility in some perfluorinated organic solvents, such as perfluoro(butyltetrahydrofuran), which allows them to be cast into thin films. These films have exceptionally good optical transparency, in addition to other perfluoropolymer properties. Teflon AF has the lowest dielectric constant known for a solid organic polymer and is highly permeable to gases. The latter property is not shared by Cytop.



Teflon AF

Cytop

Figure 2
Amorphous perfluorinated plastics.

Both amorphous perfluoropolymers are much more costly (by one to two orders of magnitude) than the other perfluoroplastics, so they are used in extremely high value-in-use applications. Examples include anti-reflective coatings, dust defocusing films ("pellicles") for semiconductor manufacture, and optical fiber cladding.

2. Elastomers

The commercially available fluoroelastomers with carbon backbones consist of three copolymer families: VDF/HFP, TFE/propylene, and a perfluoroelastomer based on TFE/PMVE (Logothetis 1994). As with the fluoroplastics, the incorporation of fluorine significantly improves stability towards chemicals, heat, and oxidation compared to most nonfluorinated analogs.

Useful thermosetting elastomers have little or no crystallinity and glass transition temperatures below normal use temperatures, so they generally consist of long, mostly linear, and highly flexible molecules. They are cross-linked (cured) to give three-dimensional networks after forming into the final product, to insure shape recovery after large deformations. Much of the challenge in the commercial development of the fluoroelastomers came from the need for efficient cross-linking chemistry on polymer chains which were designed for chemical inertness. Cross-linking must occur at the right point in processing, and must not unacceptably compromise polymer properties.

The three carbon-based fluoroelastomer families have relatively high glass transition temperatures. Thus, they become stiff and lose sealing force at relatively high lower service temperature (Table 2). Two partially fluorinated elastomers with inorganic backbones (Fig. 3), fluorosilicone and fluorophosphazene, are notable for their very low glass transition temperatures, making them useful at temperatures as low as −65 °C. However, their high temperature stability is inferior to the carbon-backbone fluoroelastomers, so a need still exists for a fluoroelastomer which can cover the full temperature range. The inorganic backbone elastomers also tend to be less solvent-resistant than the carbon-backbone structures, and they will not be discussed further here.

2.1 VDF/HFP Copolymers

Poly(vinylidene fluoride) and its copolymers containing small amounts of HFP are semicrystalline. If the HFP content is increased to 20% or more,

Table 2
Service temperature ranges of fluoroelastomers.

Type	Service temperature (°C)
VDF/HFP	−18 to 210
VDF/HFP/TFE	−12 to 230
TFE/propylene	5 to200
TFE/PMVE	0 to 260
Fluorosilicone	−65 to 175
Fluorophosphazene	−65 to 175

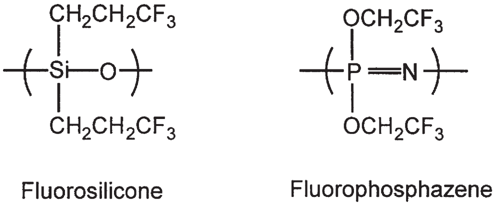


Figure 3
Fluorinated elastomers with inorganic backbones.

crystallization is prevented by the large number of –CF_3 side chains, and this system forms the basis for the most widely used fluoroelastomers. Some grades may also contain significant amounts of TFE to increase thermal and chemical stability, at some cost to low-temperature flexibility. In another variation, HFP is replaced by CTFE. The VDF/CTFE copolymers were actually the first fluoroelastomers to be developed, but are currently produced in a much smaller volume.

These partially fluorinated polymers are generally made by aqueous emulsion polymerizations in either batch or continuous systems, using conditions similar to the process for making FEP. After melt-processing into the desired shape, the polymer chains may be cross-linked by base-induced elimination of HF from the polymer backbone, followed by addition of bis-nucleophiles to the resulting electrophilic double bonds. The first cross-linking system used diamines, such as hexamethylenediamine, but the diamine cure has been largely replaced by a more stable and easily handled system which uses bis-phenols as cross-linking agent and phosphonium or ammonium salts as accelerators. In both cases, metal oxides are typically added as acid scavengers. The cross-linking mechanism has been investigated in detail (Logothetis 1989). Another curing system involves incorporation of an additional monomer (known as a cure-site monomer), typically a bromine-containing material such as bromotrifluoroethylene or 4-bromo-3,3,4,4-tetrafluorobutene. The bromine can be abstracted by peroxide-generated radicals to form a radical site on the polymer backbone. Addition of multiple polymer radicals to an added di- or trifunctional olefinic reagent generates cross-links, which tend to be somewhat more stable and rapidly formed than those generated by the bis-nucleophile cures. However, the cure site monomers add significantly to the polymer cost.

The VDF/HFP elastomers can be used at temperatures from about −20 °C to as high as 230 °C, depending on the polymer composition and cross-linking system. They are highly resistant to attack by hydrocarbons, water, acids, and ozone but swell in polar solvents such as ketones, alcohols, esters, and ethers. They are sensitive to bases such as organic amines, due to the dehydrofluorination process mentioned earlier. In common with most organic elastomers, carbon black and mineral fillers can be added to improve mechanical properties and reduce cost. Major applications are in automobiles for fuel system components and engine seals. Other applications include seals that resist corrosive environments in power plant exhausts, chemical processes, and oil wells.

2.2 TFE/Propylene Copolymers

As with ethylene, TFE copolymerizes with propylene to give a nearly alternating copolymer. The propylene-based material is amorphous, due to the random configuration of methyl groups along the polymer chain, and has a glass transition temperature of about 0 °C, so behaves as an elastomer.

TFE/propylene copolymers are prepared by aqueous emulsion polymerization, using a redox initiator at low temperatures (close to room temperature) to minimize chain transfer and branching from the propylene comonomer. The elastomer is generally cross-linked using a peroxide plus a radical trap. It exhibits good thermal stability, with an upper use temperature of about 200 °C and excellent resistance to oxidizing agents. It is more resistant to polar solvents than the VDF/HFP and is stable to amines, but swells in aromatic solvents.

2.3 Perfluoroelastomer

The VDF/HFP and TFE/propylene elastomers contain hydrogen, so they have limited resistance to very harsh environments. A perfluorinated copolymer with the chemical, thermal, and oxidative resistance of PTFE and the mechanical properties of an elastomer is produced by copolymerization of TFE with perfluoromethylvinyl ether (PMVE). The copolymer contains about 30 mol.% PMVE and is made by aqueous emulsion polymerization.

A major challenge in developing the perfluorinated elastomer was finding ways to cross-link it without compromising the stability of the final parts. In practice, this is accomplished by adding one of several termonomers to the polymer backbone at low levels (about 1 mol.%). The termonomers are usually fluorinated vinyl ethers containing a reactive functional group such as pentafluorophenoxy or nitrile. The former system is cured by reaction with a bisphenol, the latter by trimerization in the presence of a catalyst. A brominated cure site monomer allows a radical cure similar to that used for the partially fluorinated elastomers. The ultimate chemical and thermal stability of perfluoroelastomer parts depends on the cross-linking system. The nitrile system is normally considered the best, but is the most costly due to the very expensive cure site monomer.

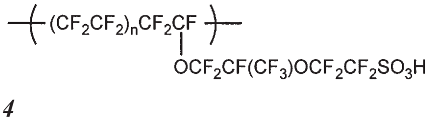
The perfluorinated elastomer is useful for long periods at temperatures up to 260 °C and for brief periods as high as 316 °C. It is inert to oxidation, chemicals, and solvents under most conditions, but can swell in some solvents depending on the cure system employed. Major applications are o-rings and seals used in highly aggressive environments such as deep oil wells and jet engines.

3. *Ion Exchange Resins*

The fluorinated ion exchange resins have a perfluorinated backbone and fluorinated side chains which terminate in sulfonic or carboxylic acid groups (Grot 1994). The structure in Fig. 4 is representative. The polymers are produced by copolymerization of TFE with the functionalized perfluorovinyl ether monomers (in their sulfonyl fluoride or methyl ester form). The resulting copolymers, which typically contain a 6–7 to 1 ratio of TFE to comonomer, can be melt-processed into films and the functional groups then hydrolyzed to their acid or salt forms.

The fluorinated ion exchange resins have several key properties. A sulfonic or carboxylic acid attached to a fluorocarbon chain is several orders of magnitude more acidic than its hydrocarbon analog. The perfluoroalkylsulfonic acid group is considered a superacid. When hydrated, the acid groups form water-swollen clusters which phase separate from the highly hydrophobic perfluoropolymer backbone, giving channels which can rapidly conduct protons (or other cations) but resist the passage of anions. Finally, the perfluorocarbon backbone provides extreme chemical and thermal stability.

The major application for the ion exchange resins is as membranes to separate the anode and cathode compartments in chloralkali cells. These cells produce caustic soda and chlorine by the electrolysis of brine. The membrane transmits sodium ions to maintain electrical neutrality in the cells, but resists the passage of chloride ions which would contaminate the caustic solution. The perfluoropolymer structure is key to a long lifetime in this extremely corrosive and oxidizing environment. The ion exchange resins have also been proposed as strong, solid acid catalysts in chemical processes, and as membranes in fuel cells.



$$-\!\left(\mathrm{(CF_2CF_2)_nCF_2}\underset{\mathrm{OCF_2CF(CF_3)OCF_2CF_2SO_3H}}{\mathrm{CF}}\right)\!-$$

Figure 4
Fluorinated ion-exchange resin.

4. *Concluding Remarks*

Fluoropolymers are made in the USA, Europe, Japan, Russia, and China. Worldwide annual production in the late 1990s was approximately 85000 tonnes for fluoroplastics and 10000 tonnes for fluoroelastomers. These are small volumes compared to many polymers such as polyethylene. However, their selling prices range from approximately \$12 $\mathrm{Kg^{-1}}$ for PTFE to hundreds, or even thousands, of dollars per kilogram for the amorphous perfluoroplastics and perfluorinated elastomer. The high prices result from costly monomers, the relatively small-scale processes, and the safety requirements needed for handling some fluoromonomers. Their unique combination of properties provides a high value in many applications, so these costs can be tolerated.

Other fluorinated polymers of commercial interest are polyacrylates and methacrylates which have fluorinated alcohol groups (Rao and Baker 1994) and perfluorinated polyethers (Scheirs 1997). The former are used as surface treatment agents to prevent soiling. The latter contain $-\mathrm{OCF_2CF_2}-$, $-\mathrm{OCF_2}-$, $-\mathrm{OCF_2CF(CF_3)}-$, and/or $-\mathrm{OCF_2CF_2CF_2O}-$ repeating units and are used as oils and greases under extremely demanding conditions.

Bibliography

Anand M, Hobbs J P, Brass I J 1994 Surface fluorination of polymers. In: Banks R E, Smart B E, Tatlow J C (eds.) *Organofluorine Chemistry. Principles and Commercial Applications*. Plenum, New York, Chap. 22

Blanchet T A 1997 Polytetrafluoroethylene. *Plast. Eng. (New York)* **41**, 981–1000

Cassidy P E, Fitch J W 1997 Hexafluoroisopropylidene-containing polymers. In: Scheirs J (ed.) *Modern Fluoropolymers*. Wiley, Chichester, UK, Chap. 8

Feiring A E 1994 Fluoroplastics. In: Banks R E, Smart B E, Tatlow J C (eds.) *Organofluorine Chemistry. Principles and Commercial Applications*. Plenum, New York, Chap. 15

Grot W G 1994 Perfluorinated ion exchange polymers and their use in research and industry. *Macromol. Symp.* **82**, 161–72

Iezzi R A 1997 Fluoropolymer coatings for architectural applications. In: Scheirs J (ed.) *Modern Fluoropolymers*. Wiley, Chichester, UK, Chap. 14

Kerbow D L 1997 Ethylene–tetrafluoroethylene copolymer resins. In: Scheirs J (ed.) *Modern Fluoropolymers*. Wiley, Chichester, UK, Chap. 15

Logothetis A L 1989 Chemistry of fluorocarbon elastomers. *Prog. Polym. Sci.* **14**, 251–96

Logothetis A L 1994 Fluoroelastomers. In: Banks R E, Smart B E, Tatlow J C (eds.) *Organofluorine Chemistry. Principles and Commercial Applications*. Plenum, New York, Chap. 16

Rao N S, Baker B E 1994 Textile finishes and fluorosurfactants. In: Banks R E, Smart B E, Tatlow J C (eds.) *Organofluorine Chemistry. Principles and Commercial Applications*. Plenum, New York, Chap. 14

Resnick P R, Buck W H 1997 Teflon®AF amorphous fluoropolymers. In: Scheirs J (ed.) *Modern Fluoropolymers*. Wiley, Chichester, UK, Chap. 22

Romack T J, DeSimone J M, Treat T A 1995 Synthesis of tetrafluoroethylene-based, nonaqueous fluoropolymers in supercritical carbon dioxide. *Macromolecules* **28**, 8429–30

Scheirs J 1997 Perfluoropolyethers (synthesis, characterization and applications). In: Scheirs J (ed.) *Modern Fluoropolymers*. Wiley, Chichester, UK, Chap. 24

Smart B E 1994 Characteristics of C–F systems. In: Banks R E, Smart B E, Tatlow J C (eds.) *Organofluorine Chemistry. Principles and Commercial Applications*. Plenum, New York, Chap. 3

Sugiyama N 1997 Perfluoropolymers obtained by cyclopolymerization and their applications. In: Scheirs J (ed.) *Modern Fluoropolymers*. Wiley, Chichester, UK, Chap. 28

Yamabe M 1994 Fluoropolymer coatings. In: Banks R E, Smart B E, Tatlow J C (eds.) *Organofluorine Chemistry. Principles and Commercial Applications*. Plenum, New York, Chap. 17

Yamabe M, Matsuo M, Miyake H 1997 Fluoropolymers and fluoroelastomers. *Plast. Eng. (New York)* **40**, 429–46

A. E. Feiring
DuPont, Wilmington, Delaware, USA

Fullerenes

The fullerene family represents a new molecular form of carbon, of remarkable interest for both its chemical and physical properties. In 1985 the highly stable molecule Buckminsterfullerene (C_{60}) was identified for the first time (Fig. 1). C_{60} is the archetypal member of the fullerene family and can be considered as the third allotrope of carbon after graphite and diamond. All fullerenes have an even number of carbon atoms arranged over the surface of a closed hollow cage. Each atom is trigonally linked to three near neighbors, with three of the four valence electrons involved in sp^2 σ-bonding. The remaining p electrons are delocalized in π-molecular orbitals covering the outside and inside surfaces of the molecule. The resulting cloud of π-electron density is similar to that which covers sheets of graphite. The σ skeleton defines a polyhedral network, consisting of 12 pentagons and m hexagons, where m, according to Euler's law, is any number, other than one (including zero). Carbon cages in which the pentagons are isolated are more stable than structures in which they abut (isolated pentagon rule). Using this rule, a family of closed carbon cages with $(20+2n)$ atoms ($n=0$, 2, 3, 4, …) is predicted to exist. The smallest member is fullerene-20, whose fully hydrogenated analogue ($C_{20}H_{20}$, dodecahedrane) has been synthesized. Several fullerenes, including chiral and achiral forms, have now been isolated, including C_{70}, C_{76}, C_{78}, C_{82}, C_{84}, …, some in several isomeric forms. Giant fullerenes with $n=240$, 560, 960, also exist and arrange in quasi-icosahedral shapes. In addition, the formation and properties of other curved forms of carbon that are structurally intermediate between the extended flat graphitic sheets and the small hollow fullerenes such as nanotubes and buckyonions have attracted considerable interest.

The simplest variant of C_{60} that can be envisaged consists of substituting one of the carbon atoms of the fullerene skeleton by an aliovalent atom, A, to afford $C_{59}A$. The first bulk synthesis of such a heterofullerene, the $C_{59}N$ azafullerene, was achieved in 1995. The introduction of the nitrogen atom in the fullerene cluster strongly perturbs the electronic and geometric character of the parent C_{60}.

1. *C_{60}: Buckminsterfullerene*

1.1 Discovery and Preparation

In 1985, experiments which aimed to simulate the physicochemical conditions in a cool red giant star with a plasma generated by a pulsed laser focused in a graphite target led to the detection by mass spectrometry of an exceptionally strong signal for C_{60}^{+}. The signal for C_{70}^{+} was also prominent. It was concluded that the C_{60} and C_{70} species were long-lived and their unexpected stability was commensurate with closed-cage fullerene structures. This was confirmed five years later when macroscopic quantities of C_{60} were isolated. It was found that the black soot-like material produced by a carbon arc, struck under argon (50–100 mbar pressure), consists of $10\pm5\%$ of fullerenes which could be extracted using solvents or by sublimation. Individual fullerenes were separable by standard chromatographic techniques (alumina/hexane).

Since the parent C_{60} could be produced and separated in large enough quantities, most research has concentrated on its physical and chemical properties together with those of its derivatives. Other higher fullerenes, like C_{70}, C_{76}, C_{78}, C_{82}, and C_{84}, are now available in large enough quantities and their study is developing rapidly.

1.2 Structure and Dynamics

The shape of C_{60} is very close to spherical (Table 1). Measurements on C_{60} single crystals established that its structure at room temperature is face-centred cubic (f.c.c.). The C_{60} molecules are orientationally disordered, forming a "plastic crystal" (they are all

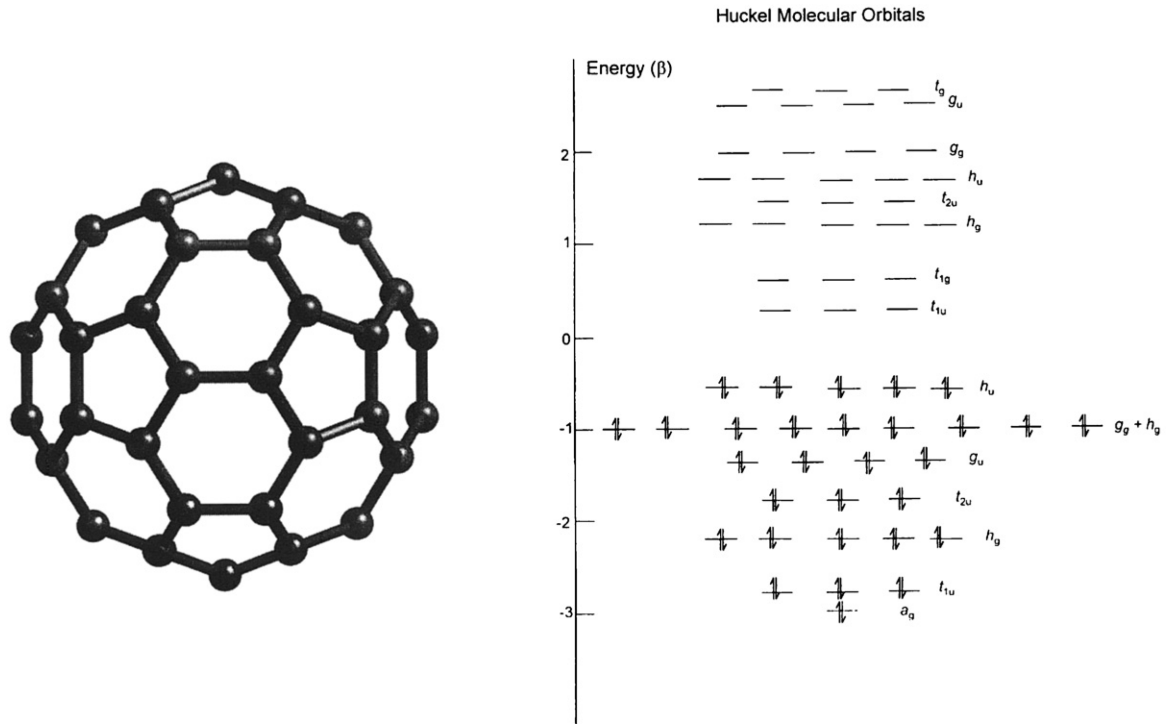


Figure 1
The C_{60} molecule (left) and its molecular orbital energy levels (right).

Table 1
Structural and electronic properties of Buckminsterfullerene, C_{60}.

Properties of C_{60}	
Molecular point symmetry	I_h
Crystal structure	face-centered cubic ($T > 260\,K$) primitive cubic ($T < 260\,K$)
Cage diameter	7.1 Å
Bond length, r(C–C)	1.404(10) Å (6:6 ring fusion) 1.448(10) Å (5:6 ring) fusion
Nearest neighbor distance	10.02 Å
Bulk modulus	18 GPa
Electron affinity	2.6–2.8 eV
Ionization energy	7.6 eV
Bandgap	1.9 eV
NMR chemical shift	142.68 ppm
IR vibrational frequencies	528, 557, 1183, 1429 cm^{-1}

spinning in different directions) with well-defined translational order and a smooth rotational potential with many shallow minima, leading to continuous small angle molecular motions. The rapid isotropic reorientation is characterized by a rotational correlation time $\tau = 9$ ps, only three times longer than that for completely unhindered (gas phase) rotation and shorter than that observed in solution.

At sufficiently low temperatures, the system adopts a crystalline structure with orientational molecular order. Below 260 K, C_{60} is primitive cubic and the C_{60} units assume well-defined orientations that minimize the intermolecular interactions. In the primitive cubic structure, electron-rich short bonds fusing two hexagons (6:6 bonds) are immediately adjacent to electron-poor pentagonal faces on nearest neighbors. In this way the electrostatic contribution to the predominantly van der Waals intermolecular interaction is optimized and the structure stabilized. In the low-temperature phase, the rotational correlation time is much longer than that at room temperature ($\tau = 64$ ns at 200 K). The dynamics are consistent with a model of jump reorientations, as the molecules shuffle between two orientations (short C–C bonds facing pentagonal or hexagonal faces). At the same time they vibrate about their equilibrium orientations. Finally, in the vicinity of 85 K, a transition to an orientational glass state occurs.

1.3 C_{60} Fullerene Derivatives

Notwithstanding its stability, which led to its original detection, C_{60} readily participates in many chemical reactions primarily acting as a soft electrophile (electron-deficient alkene) or a "radical sponge". Unlike semimetallic graphite, C_{60} is a very poor conductor with a bandgap of 1.9 eV. Reaction with good electron donors, like the alkali metals, leads to an increase in electrical conductivity of several orders of magnitude. Furthermore, some of these systems become superconducting at critical temperatures <33 K. The lowest unoccupied molecular orbital of C_{60} is triply degenerate (t_{1u}) (Fig. 1) and the stoichiometry of the superconducting alkali fullerides is A_3C_{60}. Other salts of stoichiometry A_xC_{60} ($1 \leqslant x \leqslant 11$) have also been isolated and characterized but only show insulating or metallic behavior.

C_{60} is also reduced by reaction with organic electron donors. A strong organic reducing agent such as tetrakis(dimethylamino)-ethylene (TDAE), results in the formation of a ferromagnetic solid (TDAE)C_{60} with a Curie temperature of 16 K, higher than that of any other molecular ferromagnet consisting exclusively of first-row elements.

Of considerable interest also is the possibility of encapsulating atoms (and molecules) inside the fullerene cages (incar fullerenes $i\mathrm{MC}_n$). Bulk quantities of lanthanum and scandium endohedral fullerenes $i\mathrm{MC}_{82}$ have been produced and studied by EPR spectroscopy, revealing that the metal atom has a formal charge of 3+ and C_{82} a charge of 3−; the possibility of incorporating more than one atom has also been demonstrated (e.g., $i\mathrm{Sc}_2\mathrm{C}_{82}$). The endohedral complexes may be viewed as superatoms with highly modified electronic properties, opening the way to novel materials with unique chemical and physical properties.

2. C_{70} and Higher Fullerenes

C_{70} is the second most abundant fullerene with an ellipsoidal molecular structure (point group D_{5h}). Five lines were observed in its ^{13}C NMR spectrum, implying the existence of five inequivalent carbon atoms. Even though the molecule is not spherical, at high temperature, its crystal structure is f.c.c. due to quasifree rotation. On cooling, the f.c.c. structure transforms to monoclinic via an intermediate rhombohedral phase in which there is orientational disorder only about the long ellipsoidal axis. At low temperature, the molecule is locked into an orientationally ordered phase.

Until the late 1990s it was difficult to investigate the solid higher fullerenes C_n ($n > 70$), despite their potentially interesting physical and chemical properties. This was mainly a consequence of the difficulties associated with their separation and purification. Furthermore, many higher fullerenes exist in several isomeric forms, which satisfy the isolated pentagon rule and need to be further separated. This has been achieved in 1998. C_{76} exists in two optical isomeric forms with D_2 symmetry. The enantiomeric mixtures adopt a f.c.c. structure with orientationally disordered molecules down to 100 K. In the case of C_{78}, the isolated pentagon rule excludes all but five isomers; two of these possess C_{2v} symmetry, but differ in the number of inequivalent carbon atoms. Isomer-pure solid C'_{2v} C_{78} is also f.c.c. with rotationally disordered molecules down to 100 K. C_{84} is one of the most abundant fullerene molecules after C_{60} and C_{70}. Among the 24 possible structural isomers, it is isolated as a mixture of only those with D_2 and D_{2d} symmetry in a 2:1 abundance ratio, respectively. A remarkable feature of the D_{2d} isomer is that it is the most spherical fullerene after C_{60} (aspect ratio 1.06). While the isomer mixture remains f.c.c. down to 5 K, both pure forms of D_2 and D_{2d} isomers show phase transitions to low-symmetry orientationally ordered phases, exemplifying the necessity of working on pure isomeric forms for a proper understanding of the chemical and physical properties.

3. Heterofullerenes: $C_{59}N$

The synthesis and isolation in bulk quantities of nitrogen-substituted fullerene solid offers new opportunities for the synthesis of fullerene-based materials with novel structural, electronic, and conducting properties. As a result of the trivalency of nitrogen, compared to the tetravalency of carbon, nitrogen substitution of a single carbon atom of the C_{60} skeleton leads to the very reactive azafullerene radical, $C_{59}N^{\bullet}$ that rapidly dimerizes yielding $(C_{59}N)_2$.

Theoretical calculations find that $(C_{59}N)_2$ is particularly stable, adopting a closed structure in which the nitrogens are *trans* to each other with an intermolecular C–C bond of 1.61 Å. $(C_{59}N)_2$ is a diamagnetic, insulating solid with a monoclinic structure comprising of dimers with intradimer center-to-center separation of 9.4 Å (Fig. 2). There is little mixing between the N- and C-electronic states, with strong localization of the excess electron at the nitrogen atom. Thus the chemical picture of $(C_{59}N)_2$ is consistent with the presence of a triply coordinated nitrogen atom with a lone pair in each azafullerene unit. The measured redox potentials imply that $(C_{59}N)_2$ should be easily reduced by alkali metals to afford the monomeric $C_{59}N^{6-}$ azafulleride ion which has been isolated and structurally characterized as the $K_6C_{59}N$ salt. In analogy with K_6C_{60}, it adopts a body-centred cubic (b.c.c.) structure. While band structure calculations suggest the likelihood of metallic behavior and the conductivity is much larger than that of A_6C_{60} and $(C_{59}N)_2$, magnetic measurements are consistent with the presence of unpaired spins obeying a Curie-like law. In any event, the discovery that azafullerene can give rise to a series of alkali metal intercalates opens the way to the

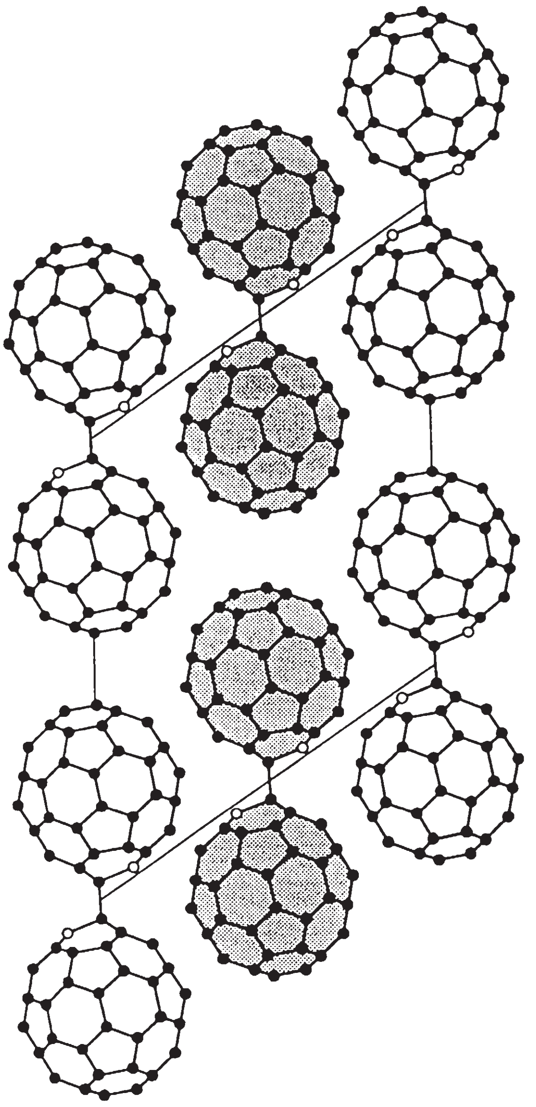


Figure 2
Basal plane projection of the crystal structure of $(C_{59}N)_2$.

synthesis of materials with electronic properties potentially as rich as those of their fullerene antecedents.

4. Fullerene Polymers

Recent advances in research on fullerene solids include the synthesis of polymeric forms of C_{60}. At ambient pressure, polymerization products form when C_{60} is exposed to a modest flux of visible or ultraviolet light between 250 and 373 K. The mechanism proposed for this photopolymerization is a [2 + 2] Diels–Alder cycloaddition, which results in the formation of four-membered carbon rings, fusing together adjacent molecules. Photopolymerization is inhibited below the ordering transition temperature as the probability of short C–C bonds coming to close intermolecular contact diminishes. C_{60} polymers also form in high-pressure/high-temperature (5 GPa, 500–800 °C) experiments. Quasi-one-dimensional bridged C_{60} structures are also encountered for the alkali fullerides, AC_{60} (A = K, Rb, Cs) and $A_2A'C_{60}$ (A = Li, Na; A' = K, Rb, Cs) and exhibit metallic behavior. At high temperature, the AC_{60} salts are f.c.c., while below 400 K the crystal structure is either orthorhombic or monoclinic with fulleride linkages along the polymer chains achieved by [2 + 2] cycloaddition in analogy with C_{60}. The center-to-center interfullerene separation along the polymer axis is

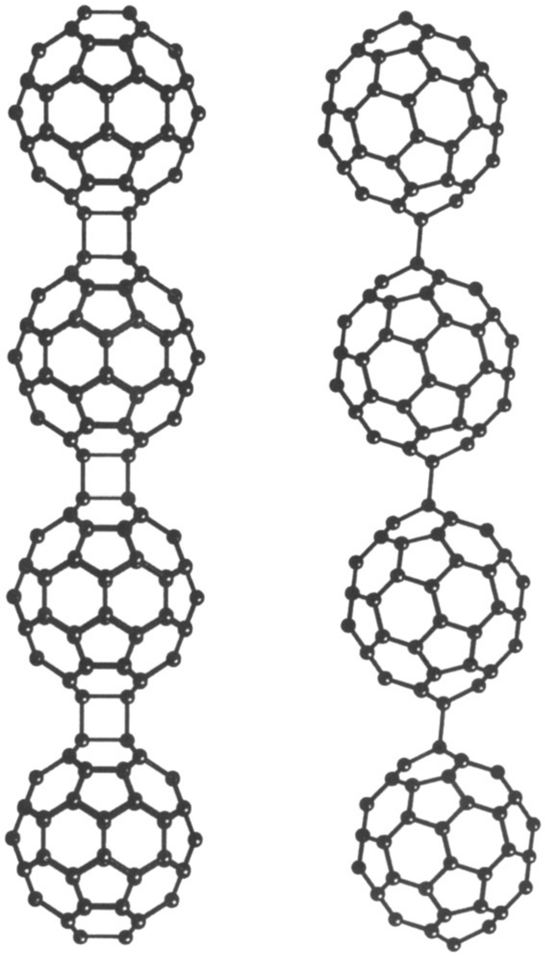


Figure 3
Polymer chains of C_{60} units, showing two different bonding geometries adopted in fulleride salts.

9.11 Å (Fig. 3, left). A completely different bonding geometry is evident in the polymeric phase of the $A_2A'C_{60}$ salts in which the fullerene molecules are connected via single C–C covalent bonds ($\sim$1.7 Å) and the nearest interfullerene distances are 9.34 Å (Fig. 3, right). A metallic polymer is also formed by Na_4C_{60} below 500 K, comprising a two-dimensional network of C_{60}^{4-} units, bridged by four single C–C bonds to its neighbors in each plane. The nearest interfullerene distances within the plane are 9.28 Å.

5. *Applications*

There are many areas in which the fullerenes and their derivatives may be uniquely useful as they exhibit features of both inorganic and organic materials. Their electronic properties suggest a possible role as conducting or semiconducting materials for batteries, transistors, and sensors. C_{60}/conducting polymer heterojunctions or composite systems have also attracted interest as new materials for organic photovoltaic cells. In these systems, C_{60} produces significant enhancement of photoconductivity because of effective photoinduced electron transfer from the conducting polymer onto C_{60}. Highly ordered thin films of C_{60} have been grown on other semiconductor substrates such as gallium arsenide opening the way to the fabrication of a new generation of advanced electronic devices. C_{60} also has unique optical properties: it is transparent to low-intensity light but nearly opaque above a critical intensity. These nonlinear optical properties suggest a role as an optical limiter in optical digital processors and for protecting optical sensors from intense light.

Other applications could come from the fact that various chemical species like metals or noble gas atoms can be attached to the surface or trapped inside (endohedral fullerenes). The properties of carbon suggest that C_{60} balls could be coated with various catalytic materials, affording a huge surface area for catalytic activity. The superconducting and ferromagnetic properties shown by some C_{60} derivatives also suggest possible applications in microelectronics devices, especially if it proves possible to synthesize air-stable, easy to handle metallic systems. Little is still known about the electronic features of the higher fullerenes but the properties of the single isolated isomers appear novel and their impact on materials science may well be exciting.

Bibliography

Andreoni W (ed.) 1996 *The Chemical Physics of Fullerenes. 10 (and 5) Years Later*. Kluwer, Dordrecht, The Netherlands

Attfield J P, Johnston R L, Kroto H W, Prassides K 1999 In: Hall N (ed.) *The Age of the Molecule*. Royal Society of Chemistry, London, pp. 181–208

Curl R F, Smalley R E 1991 Fullerenes. *Sci. Am.* **265** (4), 54

Dresselhaus M S, Dresselhaus G, Eklund P C 1996 *Science of Fullerenes and Carbon Nanotubes*. Academic Press, San Diego, CA

Gunnarsson O 1997 Superconductivity in fullerides. *Rev. Mod. Phys.* **69**, 575–606

Hirsch A 1994 *The Chemistry of the Fullerenes*. Thieme, New York

Kroto H W, Fischer J E, Cox D E (eds.) 1993 *The Fullerenes*. Pergamon, Oxford, UK

Kroto H W, Walton D R M (eds.) 1993 *The Fullerenes*. Cambridge University Press, Cambridge, UK

Prassides K, Kroto H W 1992 Fullerene physics. *Phys. World* **5**, 44–9

Prassides K (ed.) 1994 *Physics and Chemistry of the Fullerenes*. Kluwer, Dordrecht, The Netherlands

Prassides K 1997 Fullerenes. *Curr. Opin. Solid State Mater. Sci.* **2**, 433–9

Rosseinsky M J 1998 Recent developments in the chemistry and physics of metal fullerides. *Chem. Mater.* **10**, 2665–85

S. Margadonna and K. Prassides
University of Sussex, Brighton, UK

G

Glass Ceramics

Glass-ceramics are polycrystalline materials formed by the controlled crystallization of glass. Glass-ceramics can provide significant advantages over conventional glass or ceramic materials by combining the flexibility of forming and inspection of glass with improved and often unique properties in the glass-ceramic. More than US$500 million in glass-ceramic products are sold annually worldwide. These range from transparent, zero-expansion materials with excellent optical properties and thermal shock resistance to jade-like, highly crystalline materials with excellent strength and toughness. The highest volume is in cookware and tableware, architectural cladding, and oven tops and windows. Glass-ceramics are also referred to as Pyrocerams, vitrocerams, and sitalls.

The key variables in the design of a glass-ceramic are glass composition, glass-ceramic phase assemblage, and crystalline microstructure. The glass-ceramic phase assemblage (the types of crystals and the proportion of crystals to glass) is responsible for many of its physical and chemical properties, including thermal and electrical characteristics, chemical durability, and hardness. In many cases these properties are additive; for example, a phase assemblage comprising high- and low-expansion crystals has a bulk thermal expansion proportional to the amounts of each of these crystals. The nature of the crystalline microstructure (the crystal size and morphology and the textural relationship among the crystals and glass) is the key to many mechanical and optical properties, including transparency/opacity, strength, fracture toughness, and machinability. These microstructures can be quite complex and often are distinct from conventional ceramic microstructures. In many cases, the properties of the parent glass can be tailored for ease of manufacture while simultaneously tailoring those of the glass-ceramic for a particular application.

Most commercial glass-ceramic products are formed by highly automated glass-forming processes such as rolling, pressing, or casting, and subsequently converted to a crystalline product by the proper heat treatment. This heat treatment, known as ceramming, typically consists of a low-temperature hold to induce internal nucleation, followed by one or more higher-temperature holds to promote crystallization and growth of the primary crystalline phase or phases. Although some glass compositions are self-nucleating, it is more common that certain nucleating agents, such as titanates, metals, or fluorides, are added to the batch to promote phase separation and subsequent internal crystalline nucleation. Because crystallization occurs at high viscosity, article shapes are usually preserved with little or no ($<10\%$) shrinkage or deformation during the ceramming. In contrast, conventional ceramic bodies typically experience shrinkage of up to 40% during firing. Commercial products manufactured in this manner include telescope mirrors, smooth cooktops, and cookware and tableware.

Glass-ceramics can also be prepared via powder processing methods in which glass frits are sintered and crystallized. Conventional ceramic processing techniques such as spraying, tape- and slip-casting, isostatic pressing, or extrusion can be employed. Such so-called devitrifying frits are employed extensively as sealing frits for bonding glasses, ceramics, and metals. Other applications include co-fired multilayered substrates for electronic packaging, matrices for fiber-reinforced composite materials, refractory cements and corrosion-resistant coatings, bone and dental implants and prostheses, architectural panels, and honeycomb structures in heat exchangers.

1. Properties of Glass-ceramics

1.1 Thermal Properties

Many commercial glass-ceramics capitalize on their superior thermal properties, particularly ultralow thermal expansion coupled with high thermal stability and thermal shock resistance, properties that are not readily achievable in glasses or ceramics. Linear thermal expansion coefficients in the range -7.5×10^{-6}–$20\times10^{-6}\,^{\circ}\mathrm{C}^{-1}$ can be obtained. Zero or near-zero expansion materials are used in high-precision optical applications such as telescope mirror blanks as well as for oven cooktops, woodstove windows, and cookware. Glass-ceramics based on β-eucryptite solid solution display strongly negative thermal expansion (shrinking as temperature increases) and can be valuable for athermalizing precise fiber optic components. At the other extreme, high-expansion devitrifying frits are employed for sealing metals or as corrosion-resistant coatings for metals.

Glass-ceramics have high temperature resistance intermediate between that of glass and of ceramics; this property depends most on the composition and amount of residual glass in the material. Generally, glass-ceramics can operate for extended periods at temperatures of $700\,^{\circ}\mathrm{C}$ to over $1200\,^{\circ}\mathrm{C}$. Thermal conductivities of glass-ceramics are similar to those of glass and much lower than those of conventional aluminum oxide-based ceramics. They range from $0.5\,\mathrm{W m^{-1} K^{-1}}$ to $5.5\,\mathrm{W m^{-1} K^{-1}}$.

1.2 Optical Properties

Glass-ceramics may be either opaque or transparent. Their degree of transparency is a function of crystal size and birefringence, interparticle spacing, and of the difference in refractive index between the crystals and the residual glass. When the crystals are much smaller than the wavelength of light or when the crystals have low birefringence and the indices of refraction are closely matched, excellent transparency can be achieved.

Certain glass-ceramic materials also exhibit potentially useful electrooptic effects. These include glasses with microcrystallites of cadmium sulfoselenides, which show a strong nonlinear response to an electric field, as well as transparent glass-ceramics based on ferroelectric crystals such as niobates, tantalates, titanates, or lead germanates. Such crystals permit electric control of scattering and other optical properties.

1.3 Chemical Properties

The chemical durability is a function of the durability of the crystals and the residual glass. Generally, highly siliceous glass-ceramics with low-alkali residual glasses, such as glass-ceramics based on β-quartz and β-spodumene, have excellent chemical durability and corrosion resistance similar to that obtained in borosilicate glasses.

1.4 Mechanical Properties

Like glass and ceramics, glass-ceramics are brittle materials that exhibit elastic behavior up to the strain that yields breakage. Because of the nature of the crystalline microstructure, however, strength, elasticity, toughness (resistance to fracture propagation), and abrasion resistance are higher in glass-ceramics than in glass. Their strength may be augmented by techniques that impart a thin surface compressive stress to the body. These techniques induce a differential surface volume or expansion mismatch by means of ion exchange, differential densification during crystallization, or by employing a lower expansion surface glaze.

The modulus of elasticity is in the range 80–140 GPa (12×10^6–20×10^6 psi) in glass-ceramics, compared to about 70 GPa (10×10^6 psi) in glass. Abraded modulus of rupture values in glass-ceramics are in the range 50–300 MPa (7250–43 500 psi), compared to 40–70 MPa (5800–10 000 psi) in glass. Fracture toughness values are in the range 1.5–5.0 MPa $m^{1/2}$ in glass-ceramics, compared with less than 1.5 MPa $m^{1/2}$ in glass. Knoop hardness values of up to 1000 can be obtained in glass-ceramics containing particularly hard crystals such as sapphirine.

1.5 Electrical Properties

The dielectric properties of glass-ceramics strongly depend on the nature of the crystal phase and on the amount and composition of the residual glass. In general, glass-ceramics have such high resistivities that they are used as insulators. Even in relatively high-alkali glass-ceramics, alkali migration is generally limited, particularly at low temperatures, because the ions are either incorporated in the crystal phase or they reside in isolated pockets of residual glass. Nevertheless, by suitably tailoring the crystal phase and microstructure, reasonably high ionic conductivities can be achieved in certain glass-ceramics, particularly those containing lithium.

Glass-ceramic loss factors are low, generally much less than 0.01 at 1 MHz and 20 °C. Glass-ceramics comprised of cordierite or mica crystals, for example, typically provide loss factors less than 0.001. The fine-grained, homogeneous, nonporous nature of glass-ceramics also gives them high dielectric breakthrough strengths, especially compared with ceramics, allowing them to be used as high-voltage insulators or condensers.

Most glass-ceramics have low dielectric constants, typically 5–8 at 1 MHz and 20 °C. Glass-ceramics comprised primarily of network formers can have dielectric constants as low as 4, with even lower values ($K < 3$) possible in microporous glass-ceramics. Very high dielectric constants of over 1000 can be obtained from relatively depolymerized glasses with crystals of high dielectric constant, such as lead or alkaline earth titanate.

2. Glass-ceramic Systems

All commercial as well as most experimental glass-ceramics are based on silicate bulk glass compositions, although there are numerous nonsilicate and even nonoxide exceptions. Glass-ceramics can be further classified by the composition of their primary crystalline phases, which may consist of silicates, oxides, phosphates, borates, or fluorides. Examples of commercial glass-ceramics are given in Table 1.

2.1 Glass-ceramics Based on Aluminosilicate Crystals

These silicates generally consist of frameworks of silica and alumina tetrahedrons linked at all corners to form three-dimensional networks. Commercial glass-ceramics based on framework structures comprise compositions from the Li_2O–Al_2O_3–SiO_2 (LAS) and the MgO–Al_2O_3–SiO_2 (MAS) systems. The most important, and among the most widespread commercially, are glass-ceramics based on cordierite, β-spodumene, and the various structural derivatives of high (β)-quartz. These silicates are important because they possess very low bulk thermal expansion characteristics, with the consequent benefits of exceptional thermal stability and thermal shock resistance. Thus, materials based on these crystals can suffer large

Table 1
Compositions of commercial glass-ceramics (wt.%).

	Commercial applications and sources[a]						
	(A)	(B)	(C)	(D)	(E)	(F)	(G)
SiO_2	68.8	55.5	63.4	69.7	56.1	47.2	60.9
Al_2O_3	19.2	25.3	22.7	17.8	19.8	16.7	14.2
Li_2O	2.7	3.7	3.3	2.8			
MgO	1.8	1.0	NA	2.6	14.7	14.5	5.7
ZnO	1.0	1.4	1.3	1.0			
BaO	0.8		2.2				
P_2O_5		7.9	NA				
CaO					0.1		9.0
Na_2O	0.2	0.5	0.7	0.4			3.2
K_2O	0.1		NA	0.2		9.5	1.9
Fe_2O_3	0.1	0.03	NA	0.1	0.1		2.5
MnO							2.0
B_2O_3						8.5	
F						6.3	
S							0.6
TiO_2	2.7	2.3	2.7	4.7	8.9		
ZrO_2	1.8	1.9	1.5	0.1			
As_2O_3	0.8	0.5	NA	0.6	0.3		
Primary phases	β-Quartz	β-Quartz	β-Quartz	β-Spodumene	Cordierite	Fluormica cristobalite	Diopside

a (A) Transparent cookware (Visions); Corning Consumer Products Company. (B) Telescope mirrors (Zerodur); Schott Glaswerke. (C) IR transmission cooktop (Ceran); Nippon Electric Glass. (D) Cookware, hotplate tops; Corning Consumer Products Company. (E) Radomes; Corning Incorporated. (F) Machinable glass-ceramic (Macor); Corning Incorporated. (G) Gray slagsitall; Hungary.

thermal upshock or downshock without experiencing strain that can lead to rupture, a critical property in products like missile nose cones and cookware.

(a) β-Quartz and β-spodumene solid solution

Glass-ceramics in the LAS system have great commercial value for their very low thermal expansion and excellent chemical durability. These glass-ceramics are based on essentially monophase assemblages of either β-quartz or β-spodumene (keatite) solid solution, with only minor residual glass or accessory phases. Glass-ceramics containing β-quartz or β-spodumene can be made from glass of the same composition by modifying its heat treatment: β-quartz is formed by ceramming at or below 900 °C, and β-spodumene by ceramming above 1000 °C.

A mixture of ZrO_2 and TiO_2 produces highly effective nucleation of β-quartz, resulting in very small (<100 nm) crystals. This fine crystal size, coupled with low birefringence in the β-quartz phase and closely matched refractive indices in the crystals and residual glass, results in a transparent yet highly crystalline body. The combination of near-zero thermal expansion behavior with transparency, optical polishability, excellent chemical durability, and strength greater than that of glass, makes these glass-ceramics highly suitable for use as telescope mirror blanks, ring laser gyroscopes, thermally stable platforms, IR-transmitting rangetops (Fig. 1), woodstove windows, and transparent cookware.

Opaque, low-expansion glass-ceramics are obtained by ceramming these LAS materials at temperatures above 1000 °C. The β-quartz to β-spodumene transformation takes place between 900 °C and 1000 °C and is accompanied by a five- to ten-fold increase in grain size (to 1–2 mm). When TiO_2 is used as the nucleating agent, rutile development accompanies the silicate phase transformation. The combination of larger grain size with the high refractive index and birefringence of rutile gives the glass-ceramic a high degree of opacity. Secondary grain growth is sluggish, giving these materials excellent high-temperature dimensional stability. Because of their larger grain size, these glass-ceramics are stronger than those based on β-quartz, and they are also amenable to additional surface strengthening techniques. β-Spodumene glass-ceramics are fabricated via both bulk and powder sintering techniques and have found wide use as architectural sheet, cookware, bench tops, hot plate tops, heat exchangers, valve parts, ball bearings, sealing rings, and matrices for fiber-reinforced composite materials.

(b) Cordierite glass-ceramics

These glass-ceramics in the MAS system combine high strength and good thermal stability and shock

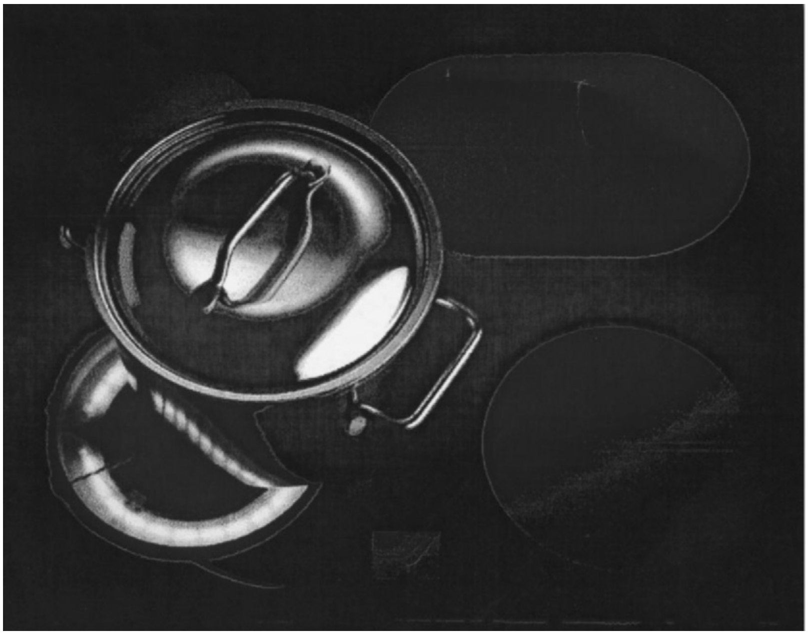


Figure 1
Radiant oven top composed of β-quartz solid solution glass-ceramic with near-zero coefficient of thermal expansion (photograph courtesy of Eurokera, SNC).

resistance with excellent dielectric properties at microwave frequencies. They are used in missile radomes and as high-performance multilayer electronic packaging.

(c) Other aluminosilicates

Transparent mullite glass-ceramics can be produced from modified binary Al_2O_3–SiO_2 glasses. When doped with ions such as Cr^{3+}, transparent mullite glass-ceramics can be made to absorb broadly in the visible while fluorescing in the near-IR, thereby making them potentially useful for luminescent solar collectors. Powder-sintered glass-ceramics based on the high-expansion aluminosilicate phase leucite ($KAlSi_2O_6$) have been evaluated for use as potential dental restorative products.

2.2 Glass-ceramics Based on Chain and Sheet Silicate Crystals

These glass-ceramics are composed of alkali and alkaline earth silicate or fluorosilicate crystals, whose crystal structures are based on single or multiple chains of silica tetrahedrons or on two-dimensional hexagonal arrays (layers) of silica and alumina tetrahedrons. Their crystal morphologies tend to reflect the anisotropy of their structures, with chain silicates typically occurring as blades or rods, and layer silicates occurring as plates. With the exception of certain lithium silicate glass-ceramics, these materials are valued most for their mechanical properties. Their microstructures of randomly oriented, highly anisotropic crystals typically give these glass-ceramics superior strength and toughness, for in order for a fracture to propagate through the material it generally will be deflected and blunted as it follows a tortuous path around or through cleavage planes of each crystal. Indeed, glass-ceramics based on chain silicate crystals have the highest toughness and body strength of any glass-ceramics.

(a) Silicate glass-ceramics

Silicate glass-ceramics include those based on lithium metasilicate (Li_2SiO_3), lithium disilicate ($Li_2Si_2O_5$), diopside ($CaMgSi_2O_6$), enstatite ($MgSiO_3$), and wollastonite ($CaSiO_3$). There are two groups of commercially important lithium silicate glass-ceramics. One group yields high-expansion lithium disilicate glass-ceramics that match the thermal expansion of several nickel-based superalloys and are used in a variety of high-strength hermetic seals, connectors, and feedthroughs. Related materials based on a microstructure of fine-grained lithium disilicate crystals with dispersed nodules of quartz crystals have been extensively evaluated for use as magnetic disk substrates for computer hard drives.

The second group is nucleated with colloidal silver, gold, or copper, which in turn are photosensitively nucleated. By suitably masking the glass and then irradiating with UV light, it is possible to nucleate and crystallize only selected areas. The crystallized portion consists of dendritic lithium metasilicate crystals, which are much more soluble in dilute hydrofluoric acid than is the glass. The crystals can thus be etched away, leaving the uncrystallized (masked) portion intact. The resulting photoetched glass can then be flood-exposed to UV rays and heat treated at higher temperature, producing the stable lithium disilicate and quartz phases. The resulting glass-ceramic is strong, tough, and faithfully replicates the original photoetched pattern. These chemically machined materials have been used as fluidic devices, lens arrays, magnetic recording head pads, and charged plates for inkjet printing.

Blast furnace slags, with added sand and clay, have been used in Eastern Europe since before the 1970s to manufacture inexpensive nonalkaline glass-ceramics called slagsitall. The primary crystalline phases are wollastonite ($CaSiO_3$) and diopside ($CaMgSi_2O_6$) in a matrix of aluminosilicate glass. Metal sulfide particles serve as nucleating agents. The chief attributes of these materials are high hardness, good to excellent wear and corrosion resistance, and low cost. Slagsitall materials have found wide use in the construction, chemical, and petrochemical industries. Applications include abrasion- and chemical-resistant floor and wall tiles, industrial machinery parts, chimneys, plungers, parts for chemical parts and reactors, grinding media, and coatings for electrolysis baths.

Translucent architectural panels of wollastonite glass-ceramics are manufactured by Nippon Electric Glass and sold under the trade name Neopariés. A sintered glass-ceramic with about 40% crystallinity, this material can be manufactured in flat or bent shapes by molding during heat treatment. It has a texture similar to that of marble, but with greater strength (50 MPa) and durability than granite or marble. Neopariés is used as a construction material for flooring and exterior and interior cladding.

(b) Fluorosilicate glass-ceramics

These glass-ceramics, like those based on the simple silicates, are primarily valued for their mechanical properties. Glass-ceramics based on the chain silicate potassium fluorrichterite ($KNaCaMg_5Si_8O_{22}F_2$) have a microstructure consisting of tightly interlocked, fine-grained, rod-shaped amphibole crystals. These glass-ceramics have good chemical durability, are usable in microwave ovens, and, when glazed, resemble bone china in their gloss and translucency. Richterite glass-ceramics are manufactured for use as high-performance institutional tableware.

Machinable glass-ceramics based on sheet silicates of the fluorine-mica family have unique microstructures composed of interlocked, randomly oriented platy mica crystals. Because micas can be easily delaminated along their cleavage planes, fractures propagate readily along these planes but not along other crystallographic planes, thereby enabling these glass-ceramics to be readily machined. Machinable mica glass-ceramics are employed in many applications, including high-vacuum components and hermetic joints, precision dielectric insulators and components, seismograph bobbins, sample holders for field ion microscopes, boundary retainers for the Space Shuttle, gamma-ray telescope frames, and dental restorations.

The inherent strength and machinability of mica-based materials have extended the growing field of biomaterials. Biomaterials for bone implants demonstrate biocompatibility—are well tolerated by the body—and may even offer bioactivity, the ability of the biomaterial to bond with the hard tissue (bones) of the body. Several strong, machinable, and biocompatible mica- and mica/fluorapatite glass-ceramics have been widely studied as bone implants.

2.3 Glass-ceramics Based on Nonsilicate Crystals

(a) Oxides

Glass-ceramics consisting of various oxide crystals in a matrix of siliceous residual glass offer properties not available with more common silicate crystals. In particular, glass-ceramics based on spinels and perovskites can be quite refractory and can yield useful optical, mechanical, and electrical properties. Possible applications for spinel glass-ceramics include solar collector panels, liquid crystal display screens, high-temperature lamp envelopes, and magnetic disk substrates. Glass-ceramics based on perovskite crystals are characterized by their unusual dielectric and electrooptic properties. Examples include highly crystalline niobate glass-ceramics that exhibit nonlinear optical properties, as well as titanate, niobate, and tantalate glass-ceramics with very high dielectric constants.

(b) Phosphates

Many phosphates claim unique material advantages over silicates that make them worth the higher material costs for certain applications. Glass-ceramics containing the calcium orthophosphate apatite, for example, have demonstrated good biocompatibility and, in many cases, even bioactivity, making them useful as bone implants and prostheses. Mixed phosphate–silicate glass-ceramics based on fluorapatite and wollastonite, fabricated using conventional powder processing techniques, are bioactive and have good flexural strength of up to 200 MPa. The aforementioned combination of fluorapatite with phlogopite mica provides bioactivity as well as machinability. Cast glass-ceramics with microstructures comprising interlocking needles of fluorapatite and mullite provide high fracture toughness, with K_{1c} values greater than $3\,\mathrm{MPa\,m^{1/2}}$.

Certain glasses in the B_2O_3–P_2O_5–SiO_2 system melted under reducing conditions can yield a unique microfoamed material on heat treatment. These materials consist of a matrix of BPO_4 glass-ceramic filled with uniformly dispersed, isolated 1–10 mm hydrogen-filled bubbles. The hydrogen evolves on ceramming, most likely owing to a redox reaction involving phosphite and hydroxyl ions. These materials, with d.c. resistivity of $10^{16}\,\Omega\,\mathrm{cm}$ at 250 °C, dielectric constants as low as 2, and densities as low as $0.5\,\mathrm{g\,cm^{-3}}$, have potential application in electronic packaging.

A novel class of microporous glass-ceramics composed of skeletons of two types of titanium phosphate crystals has been prepared by chemical etching methods analogous to those used for Vycor glasses. These materials offer promise for applications including oxygen and humidity sensors, immobilized enzyme supports, and bacteriostatic materials.

(c) Fluorides and borates

Transparent oxyfluoride glass-ceramics, consisting of fluoride nanocrystals dispersed throughout a continuous silicate glass, have been shown to combine the optical advantages of rare-earth-doped fluoride crystals with the ease of forming and handling of conventional oxide glasses. Fluorescence and lifetime measurements indicate that these materials can be superior to fluoride glasses both for Er^{3+} amplifiers because of greater width and gain flatness of the 1530 nm emission band and for 1300 nm Pr^{3+} amplifiers because of their higher quantum efficiency.

Transparent glass-ceramics based on submicrometer, spherical crystallites of β-BaB_2O_4 (BBO) in a glass of the same composition demonstrate second harmonic generation to UV wavelengths; such materials offer promise as potential optical components. Other aluminoborate glass-ceramics may be useful as low-expansion sealing frits.

3. Conclusions

Glass-ceramics have been used for many years in applications ranging from cookware and missile radomes to telescope mirrors, bone prostheses, and IR-transmitting cooktops. Future applications are likely to capitalize on highly specialized properties for the transmission, display, and storage of information.

Bibliography

Beall G H 1992 Design and properties of glass-ceramics. *Ann. Rev. Mater. Sci.* **22**, 91–119

Beall G H, Duke D A 1969 Transparent glass-ceramics. *J. Mater. Sci.* **4**, 340–52

Beall G H, Pinckney L R 1999 Nanophase glass-ceramics. *J. Am. Ceram. Soc.* **82**, 5–16

Berezhnoi A I 1970 *Glass-Ceramics and Photo-Sitalls*. Plenum, New York

Dejneka M J 1998 Transparent oxyfluoride glass-ceramics. *MRS Bull.* **11**, 57–62

Hench L L 1998 Bioceramics. *J. Am. Ceram. Soc.* **81**, 1705–28

Höland W 1997 Biocompatible and bioactive glass-ceramics—state of the art and new directions. *J. Non-Cryst. Solids* **219**, 192–7

Lewis M H (ed.) 1989 *Glasses and Glass-Ceramics*. Chapman & Hall, London

McHale A E 1992 Engineering properties of glass-ceramics. In: *Engineered Materials Handbook: Vol. 4. Ceramics and Glasses*. ASM International, Materials Park, OH, pp. 870–8

McMillan P W 1979 *Glass-Ceramics*, 2nd edn Academic Press, New York

Petzoldt J, Pannhorst W 1991 Chemistry and structure of glass-ceramic materials for high precision optical applications. *J. Non-Cryst. Solids* **129**, 191–8

Pinckney L R 1994, 4th edn. *Glass-ceramics. Kirk-Othmer Encyclopedia of Chemical Technology*, Vol. 12, pp. 627–44

Strnad Z 1986 *Glass-Ceramic Materials, Glass Science and Technology*. Elsevier, Amsterdam, Vol. 8

Vogel W, Höland W 1990 Development, structure, properties, and application of glass-ceramics for medicine. *J. Non-Cryst. Solids* **123**, 349–53

L. R. Pinckney

Corning Incorporated, Corning, New York, USA

H

High-temperature Stable Polymers

One possible definition for high temperature (HT) stable thermoplastic polymers includes materials which: allow thermoplastic processing; possess a continuous service temperature (CST) above 150 °C; are chemical and radiation resistant; and exhibit a low flammability or are even self-extinguishing. The last point is particularly important for applications in aircraft, building/construction, and electric/telecommunication industries. Thermal degradation resistance of the melt is essential for thermoplastic processing since typically amorphous polymers are processed about 130–150 °C above T_g (glass transition temperature). A semicrystalline polymer normally requires processing temperatures in the range 30–50 °C above its melting temperature (T_m).

Structural concepts underlying thermoplastic HT polymers are based on fully aromatic building blocks linked via flexible segments. In particular the incorporation of links such as oxygen, keto, and sulfone groups increases processability while maintaining acceptable temperature stability. Most commercial grades of HT thermoplastics are reinforced with fillers to further improve temperature stability and mechanical properties at higher temperatures. This article covers important amorphous HT polymers, (i.e., polysulfones (PSU) and polyetherimides), semicrystalline polymers, (i.e., polyphenylenesulfide and polyaryletherketones), and thermotropic liquid crystalline polyesters. In Table 1 the most important thermoplastic HT stable polymers are listed, completed with T_g, T_m, and CST data. Some representative trade names and suppliers are given. In addition, a selection of nonthermoplastic commercially available HT polymer products, i.e., polyimides and polybenzimidazoles, are listed. Thermosetting materials are not included in this article. All annual production figures represent the market around 1998/99 (Sparenberg *et al.* 1999). More information about other HT polymers with no or little commercial importance can be found in Sillon and Rabilloud (1995), Meador (1998), and *Fluorine-containing Polymers.*

1. *Polyarylethersulfones*

In 1998, PSU polyethersulfone (PES), and polyphenylethersulfone (PPSU) contributed more than 2.0×10^4 tyr^{-1} to the worldwide market of amorphous HT thermoplastics. Important properties are: good mechanical behavior between −50 °C and 180 °C; high dimensional stability; transparency; and self-extinguishing. However, these polymers suffer from poor UV light stability, and thus must be filled for outdoor uses, as well as decreased notch impact strength, which is improved for PPSU. Typical applications are found for domestic household goods, in food and electronic industries, and for medical instruments, which must withstand autoclave sterilization temperatures and conditions. PES is preferred in the automobile industry, since its chemical resistance is much better compared with other PSUs. PES shows excellent adhesion to metals resulting in typical applications as headlight reflectors or molded interconnected devices (MIDs).

2. *Polyetherimides*

Polyetherimide (PEI) is an amorphous HT thermoplastic manufactured by GE Plastics with an annual production of 1.2×10^4 t. With respect to its property profile, PEI competes directly with PES and shares the same application market. However, it is intrinsically more yellow and less stable towards bases. Advantageously it has excellent adhesion to glass fibers, which results in a linear relationship between flex modulus and tensile strength with the glass fiber content up to 30 wt.%. Newer developments by GE Plastics deal with improved PEI blends with reduced melt flow viscosity and enhanced transparency.

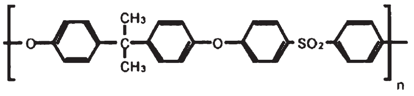
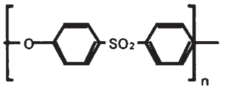
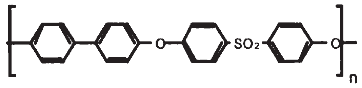
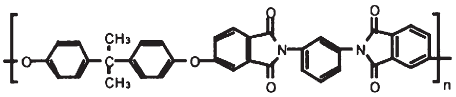
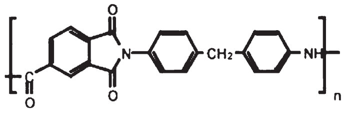
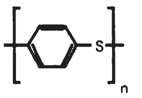
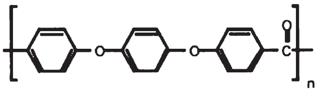
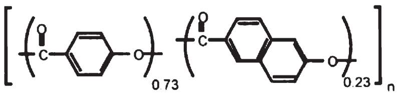
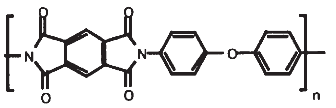
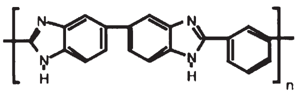
3. *Polyamidimides*

Polyamide (PAI) is manufactured by BP Amoco but also distributed by DSM Engineering Plastics. It is commercialized as melt processible pellets with lower molecular weight to reduce melt viscosity and filled with graphite, glass fibers, or polytetrafluoroethylene (PTFE). Some types are only available as semifinished parts in the form of rods, plates, or bars. Its benefits are: excellent mechanical properties in a broad temperature range of −190 °C to 260 °C; low coefficient of thermal expansion (CTE); high resistance towards oxygen, UV, and higher energy radiation; and good metallization behavior. However, in particular injection molding at high plasticizing and mold temperatures is difficult and modifications may be necessary. Typical end uses are found in the electronic and semiconductor industries and for other applications such as bearings, pistons, and seals.

4. *Polyphenylenesulfide*

Worldwide production of polyphenylsulfide (PPS) is 3.5×10^4 tyr^{-1}. PPS is a semicrystalline polymer and is distributed as linear and cured cross-linked grades. The degree of crystallinity can be as high as 60% and is extremely dependent on the thermal history. Due to

Table 1
Common high temperature stable thermoplastic and selected nonthermoplastic polymers including some imporatnt thermal properties and representative trade names.

Polymer	Abbrev.[a]	Structure	T_g (°C)	T_m (°C)	CST[b] (°C)	Trade names
Thermoplastic						
Poly-sulfone	PSU		187		160	*Ultrason® S* (BASF) *Udel®* (BP Amoco)
Poly-ethersulfone	PES		225		190	*Ultrason® E* (BASF)
Polyphenyl-ethersulfone	PPSU		221		149	*Radel® R* (BP Amoco)
Poly-etherimide	PEI		215		180	*Ultem®* (GE Plastics)
Poly-amidimide	PAI		275		260	*Torlon®* (BP Amoco)
Poly-phenylene-sulfide	PPS		90	285	220	*Fortron®* (Ticona) *Supec®* (GE Plastics) *Ryton®* (Phillips Pet.)
Polyether-etherketone	PEEK		143	343	250	*Victrex® PEEK* (Victrex plc.)
Liquid crystalline polyester	LCP		n/a	280	220	*Vectra® A* (Ticona)
Non-thermoplastic						
Polyimde	PI		360		250	*Vespel®* (DuPont)[c]
Polybenz-imidazole	PBI		425		310	*Celazole®* (DSM Eng.)[c]
Polytetra-fluoro-ethylene	PTFE	$\left[\mathrm{CF_2{-}CF_2}\right]_n$	127[d]	327	250	*Teflon®* (DuPont)

a According to ISO/DIN 1043-1. b Continuous service temperature. c Only available as semifinished parts. d Alpha glass transition.

its low T_g PPS is often distributed as a highly filled compound. Besides good thermal and mechanical properties, such as low creep, PPS shows excellent chemical resistance towards hot fuels and oil, chlorinated hydrocarbons, bases, and dilute acids and is therefore superior to PSUs. These properties determine the final uses and increasingly PPS is used as replacement for glass fiber-reinforced polyamides in the electronic and food industries, and in automobiles as fuel pumps and other under-the-hood parts. PPS can also be spun into fibers and processed into blow-molded parts.

5. *Polyaryletherketones*

Polyaryletherketones originally included several polymers abbreviated as PEKEKK, PEK, PEEKK, or PEEK, indicating a sequence of ether (E) or ketone (K) linkages. However, when BASF AG withdrew its product Ultrapek (PEKEKK) from the market, Victrex PEEK became the only commercially available polyaryletherketone with a production of more than 1100 tyr^{-1}. This class of semicrystalline HT polymer develops a high degree of crystallinity under normal processing conditions and is commonly processed for bulk parts via injection molding or compression molding. PEEK exhibits excellent solvent resistance at room temperature. Only strong acidic media can protonate the carbonyl groups and damage the polymer. Other outstanding features are almost constant electric and dielectric properties up to 200 °C, accompanied by low water absorption. Applications are found predominately in the electronic, telecommunication and semiconductor industries, but also in the aircraft, automobile, food, and textile business. For medical instruments and purposes, PEEK is selected due to its biocombatibility, FDA approval, and radiation and hydrolysis resistance.

6. *Liquid Crystalline Polyesters*

Thermotropic liquid crystalline polyesters (LCP) are manufactured by several companies, and the materials Vectra (Ticona), Xydar (BP Amoco), and Zenite (DuPont) alone contribute to almost 75% of the annual world production of around 1.9×10^4 t. The special feature which distinguishes LCP from other HT thermoplastics is their liquid crystalline order in the melt. This is important for injection molding processing, since increasing shear both reduces melt viscosity and increases the strength accompanied by low equipment wear and short processing cycles. Due to these advantages, LCP can be easily processed into extremely thin and complex small parts down to 100 μm thickness with high precision without sacrificing reproducibility and dimensional stability. Hence, the main application of LCP are found in the electronic and telecommunication industries, e.g., lamp sockets, cellular phone parts, and MIDs.

7. *Nonthermoplastic High Temperature Stable Polymers*

Polyimides are historically processed from solution in a two-step method. The two examples included in Table 1, Vespel SP and Celazole, cannot be processed via common thermoplastic techniques, but are sold as semifinished parts shaped as rods, rings, plates, or tubular bars only. Despite their high prices, the unusually high CST and enhanced tribological performance opens applications in high-tech industries such as semiconductor, aircraft, and aerospace. Vespel SP is modified with graphite, PTFE, or MoS_2 as well as glass or aramid fibers. The lubricating additives already suggest the final use as bearing components with low abrasion and excellent frictional behavior. Celazole is offered in an uncompounded grade only. Thermoplastic polyether- or polyamidimides based on flexible monomers have been developed on a small scale. PTFE (see *Fluorine-containing Polymers*) classifies also as HT stable polymer and yet cannot be processed by thermoplastic processing techniques. It is available in a wide variety of parts, sheets/films, and powders. Other HT stable polymers, e.g., Kevlar or Nomex, are processed from solution and are marketed as fibers, woven fabrics, or pulp. Their major use is based on high strength fibers in bulletproof vests, fiber reinforcement in composites, security apparel, and as high temperature materials in asbestos-free brake pads.

8. *Concluding Remarks*

HT thermoplastics have a steadily increasing market and further growth is predicted. Their higher CST, convenient processing by injection molding techniques, and meanwhile competitive pricing allow for HT thermoplastics to successfully replace traditional engineering thermoplastics and increasingly more thermosetting materials. Particularly in the electronic industry the advance of molded interconnected devices and injection molded circuit boards will require the use of HT thermoplastics with high dimensional stability, thermal resistance during soldering, effective metallization, and recyclability.

Bibliography

Domininghaus H 1998 *Die Kunststoffe und ihre Eigenschaften*, 5th edn. Springer, Berlin

Meador M A 1998 Recent advances in the development of processible high-temperature polymers. *Annu. Rev. Mater. Sci.* **28**, 599–630

Münstedt H, Streib J 1997 Hochtemperaturbeständige Thermoplaste. In: Schaumburg H (ed.) *Polymere*. Teubner, Stuttgart Chap. 5

Parker D, Bussink J, van de Grampel H T, Wheatley G W, Dorf E U, Ostlinning E, Reinking K 1992 Polymers, high-temperature. In: Elvers B, Hawkins S, Schulz G (eds.) *Ullmann's Encyclopedia of Industrial Chemistry*. 5th edn. VCH, Weinheim, Germany, Vol. A21, pp. 449–72

Rubin I I (ed.) 1990 *Handbook of Plastic Materials and Technology*. Wiley, New York

Sillon B, Rabilloud G 1995 Heterocyclic polymers with high glass transitions. In: Ebdon J R, Eastmond G C (eds.) *New Methods of Polymer Synthesis*. Blackie, London, Vol. 2 Chap. 7

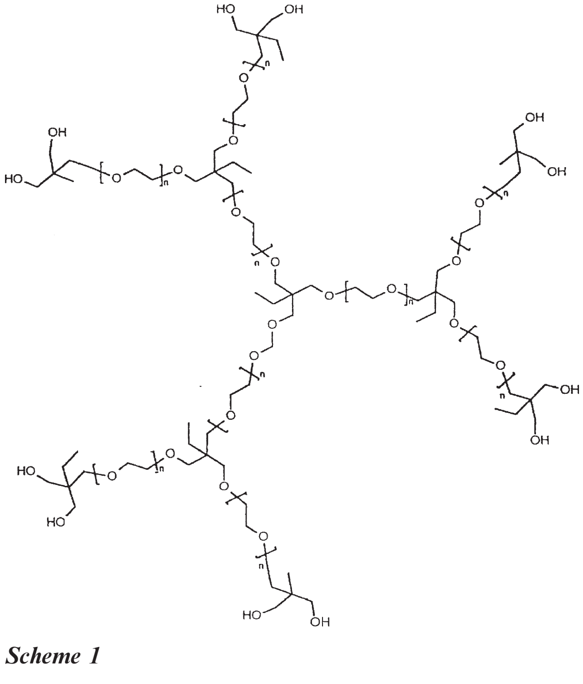
Sparenberg B, Eichenauer U, Radden P, Reimer W, Weidig R, Puyenbroek R 1999 *Hoch leistungs Kunststoffe* **89** (10), 136–60

R. Giesa and H.-W. Schmidt
University of Bayreuth, Germany

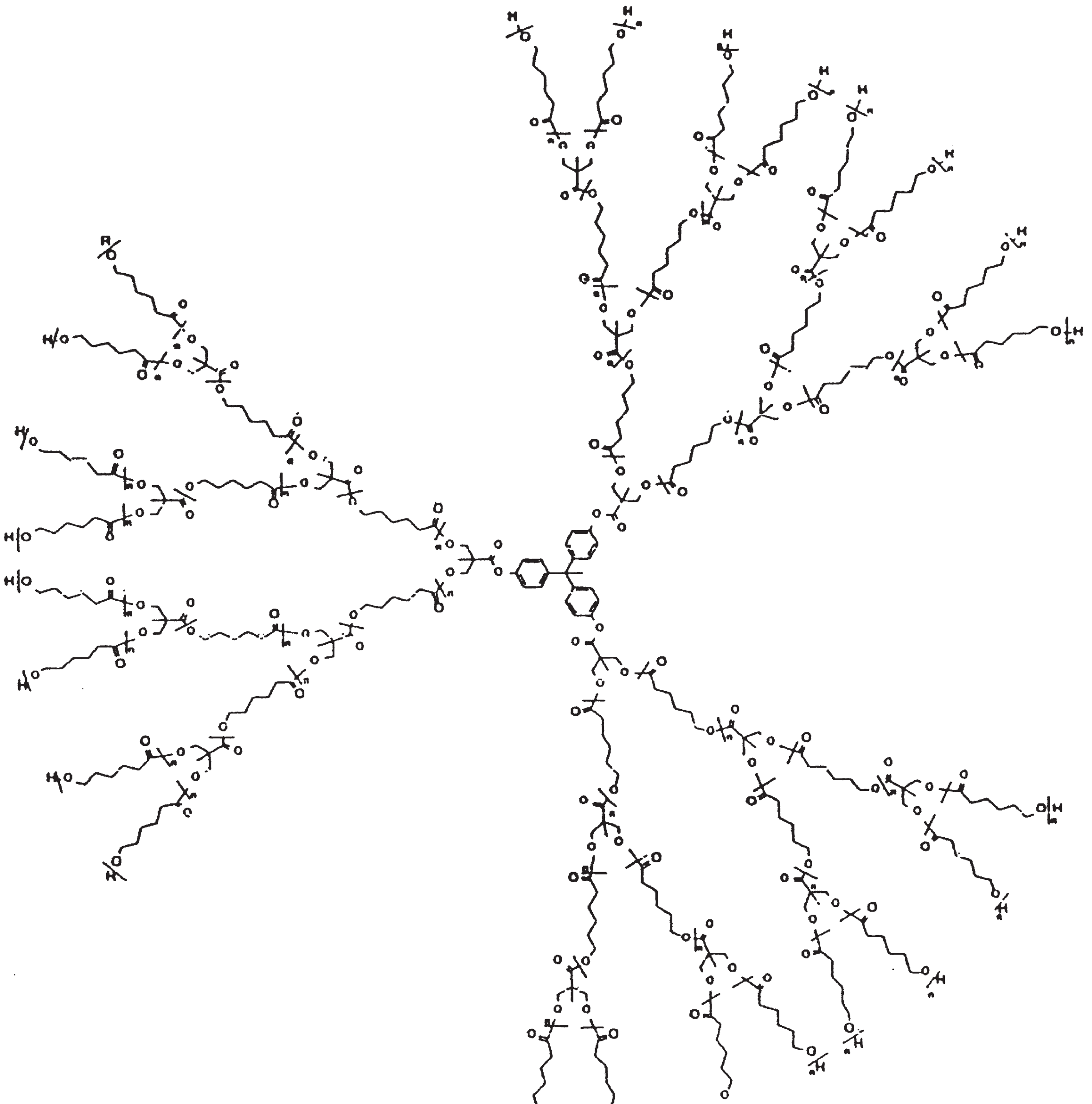
Hybrid Dendrimer Star-like Polymers

The introduction of controlled branching as a means of modifying the properties of synthetic polymers has become a challenging area of both science and technology (Frechet 1994, Webster 1994) since it has become apparent that rheological, mechanical, and solution properties correlate closely to the type and degree of branching (James 1996). One of the most important examples is polyethylene which can be obtained as either linear low density polyethylene (LDPE), high density polyethylene (HDPE), or highly branched/amorphous polyethylene (Johnson *et al.* 1995). Although each of these forms of polyethylene has the same skeletal structure, it has been well established that the different branching extents has a pronounced effect on the degree of crystallization and subsequent mechanical properties. In addition, star and graft polymers are examples of branched systems that are established in many areas of technologies as rheological control agents and emulsifiers (Roovers *et al.* 1998, Ueda *et al.* 1998). Although it has been difficult to compare star polymers with their linear analogs, it has been shown that their usefulness is based on their low hydrodynamic volume and intrinsic viscosity. Over the last decade, dendrimers and hyperbranched polymers, two other classes of highly branched polymers, have emerged and achieved commercial status (Frey *et al.* 1998, Percec *et al.* 1998). The interest in dendritic macromolecules is varied, but like star-shaped polymers, it is based on the fact that they are inherently different to their linear analogs, and it is this difference that leads to many of the unique solution and physical properties (Hawker *et al.* 1997, Mio *et al.* 1998). Although dendritic macromolecules have many interesting features and are being surveyed for many niche applications, they have extremely poor mechanical properties, owing to the absence of chain entanglement, and require many steps to build modest MWs.

To overcome these deficiencies caused by a lack of significant entanglements, hybrid dendritic linear copolymers have emerged as a novel class of polymers which exhibit properties characteristic of both dendrimers and linear polymers. One of the most well studied of these hybrid structures are dendrimer-like star polymers (Angot *et al.* 1998, Trollsås and Hedrick 1998). These polymers are characterized by having generations of high MW polymers between periodic branching junctures emanating from a central core. Using the divergent growth approach with a combination of living/controlled polymerization procedures and quantitative organic transformations, MWs as high as 250 000 g mol^{-1} are easily obtained. The first example of a dendrimer-like star polymer was reported by Six and Gnanou (1995) using anionically prepared poly(ethylene oxide) as the macromolecular building block. Two generations of poly(ethylene oxide) with narrow polydispersities were obtained (Scheme 1). However, the majority of the reports of dendrimer-like star polymers have been based on aliphatic polyesters. A dendrimer-like star polymer based on poly(ε-caprolactone) and a dendritic initiator and branching junctures based on 2,2′-bis(hydroxymethyl) propionic acid (bisMPA) and its derivatives are shown in Scheme 2 (Trollsås and



Scheme 1



Scheme 2

Hedrick 1998). Some general characteristics of dendrimer-like star polymers are given in Table 1. Considerable flexibility exists in the construction of this architecture, as the generation of the dendrimer-initiating core, average MW of the polymer, type of polymer, generation or functionality of the branching juncture between polymer generations, and polymer generation can be varied to obtain complex architectures that resemble the most advanced dendrimers. However, dendrimer-like star polymers, as will be described in this article, have features characteristic of dendrimers and star-shaped macromolecules including abundant chain end functionality, three-dimensional shape, and mechanical properties similar to their linear analogs. Moreover, the versatility of the synthetic procedure to dendrimer-like star polymers has permitted variations in the molecular structure between generations. This has enabled the preparation of both amorphous and semicrystalline morphologies, materials which undergo microphase

Table 1
Characteristics of dendrimer-like star polymers.

Sample entry	Target DP	Monomer	M_n (^{1}H-NMR)	M_n (SEC)	MW/M_n	T_m (°C), ΔH (Jg^{-1})
G1-5	5	ε-CL	4 500		1.07	12.3, 48.1
G2-5	5	ε-CL	12 700	18 800	1.06	21.1, 49.3
G3-5	5	ε-CL	33 100	55 400	1.30	23.5, 45.0
G1-10	10	ε-CL	9 200	14 900	1.13	39.1, 59.7
G2-10	10	ε-CL	23 600	38 200	1.09	38.8, 60.7
G3-10	10	ε-CL	53 300	75 000	1.30	36.5, 59.2
G1-15	15	ε-CL	10 400	17 400	1.09	43.0, 66.2
G2-15	15	ε-CL	49 400	49 400	1.11	43.0, 61.0
G3-15	15	ε-CL	80 000	80 000	1.09	40.6, 61.6
G1-20	20	ε-CL	24 400	24 400	1.11	44.4, 69.7
G2-20	20	ε-CL	74 700	74 700	1.22	46.2, 57.1
G3-20	20	ε-CL	102 000	102 000	1.21	46.8, 63.2
G1-20	20	*L*-Lactide		23 500	1.04	147.0, 30.3
G2-20	20	*L*-Lactide		58 700	1.18	148.0, 22.7
G3-20	20	*L*-Lactide		101 100	1.26	143.0, 18.0
G1-20	20	*Rac*-Lactide		16 600	1.07	
G2-20	20	*Rac*-Lactide		36 300	1.13	
G3-20	20	*Rac*-Lactide		76 700	1.39	

separation and materials which are stimuli responsive (i.e., micelle-mimics).

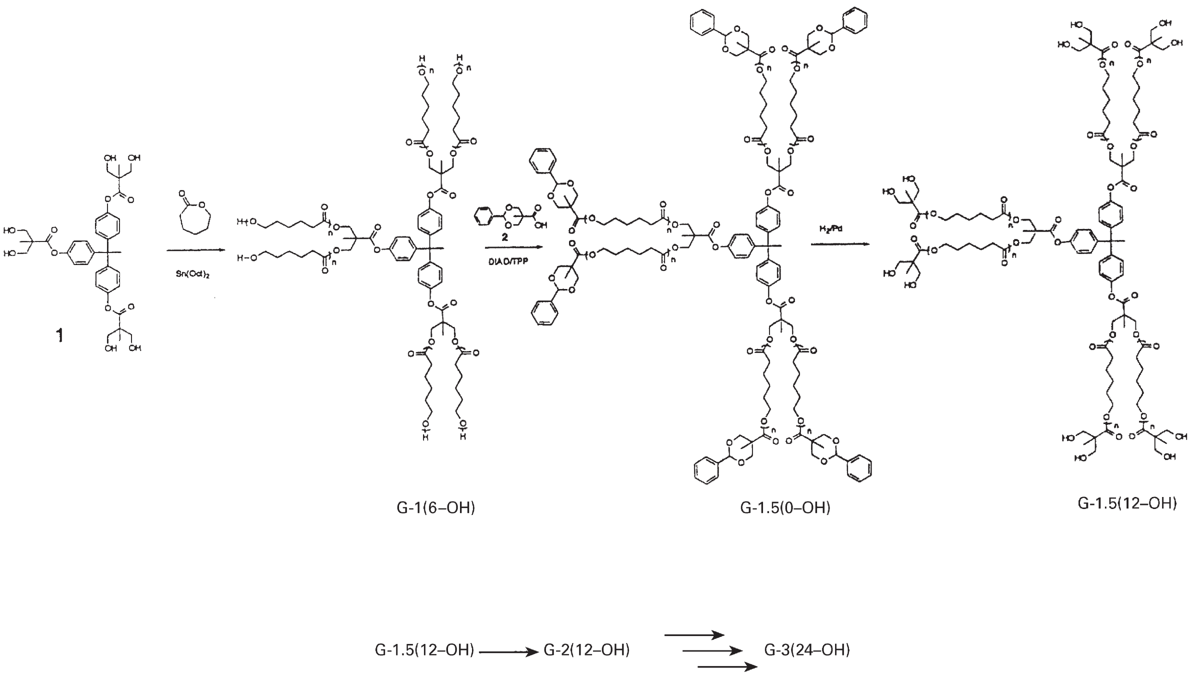
1. *Synthesis*

The majority of the synthetic routes to dendrimer-like star polymers have focused on the ring-opening polymerization (ROP) of either ethylene oxide or cyclic esters and diesters. For example, the dendrimer-like star poly(ε-caprolactones) were prepared by the ROP of ε-caprolactone initiated from bis(hydroxymethyl) derivatives in the presence of stannous-2-ethylhexanoate (Sn(Oct)$_2$) (Trollsås *et al.* 1997, Hecht *et al.* 1999). Polymerization of poly(ε-caprolactone) in this manner produced remarkably narrowly dispersed products with accurate control of MW, as determined from the monomer to initiator ratio. Furthermore, near quantitative conversion of monomer to polymer was observed, and the MW attainment was linear with conversion, suggesting living character.

A typical synthesis of three generations of dendrimer-like star poly(ε-caprolactone) is shown in Scheme 3 (Trollsås and Hedrick 1998). Polymerization of ε-caprolactone was initiated from a hexa hydroxyl functionalized first generation dendrimer of bisMPA using Sn(Oct)$_2$ in bulk conditions to give a six arm star-shaped polymer, **G1**. The hydroxyl functional arms of **G1** were then esterified with a benzylidene protected bisMPA using Mitsunobo conditions followed by deprotection by hydrogenolysis to generate **G1.5**, a six-arm star polymer with 12 hydroxyl groups. This polymer can be used as a "macroinitiator" for the ROP of ε-caprolactone to give the second generation dendrimer-like star polymer **G2** with 12 arms and 12 chain-end hydroxyl groups. Another iteration of the same procedure generated the third generation dendrimer-like star polymer, **G3**, having 24 arms (Trollsås *et al.* 2000). Several series of dendrimer-like star polymers were also prepared from ε-caprolactone, L-lactide, and the racemate of L- and D-lactide(*rac*-lactide) (Althoff *et al.* 1998). Figure 1 shows the size exclusion chromatograms (SEC) for the second generation polymers which clearly shows that the size of the dendrimer-like star polymer can be altered simply by varing the DP.

2. *Characterization*

Figure 2 shows the SEC chromatograms for three generations of dendrimer-like star polymers (DP = 20), which show narrowly dispersed products are obtained with higher elution volumes with increasing generation number (Trollsås *et al.* 1998). However, the shift in the hydrodynamic volume between the second and the third generations does not



Scheme 3

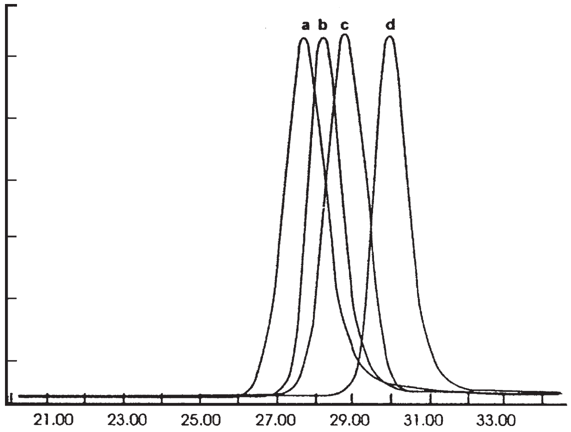


Figure 1
SEC chromatograms of the second generation G-2 dendrimer-like star polymers of different average DPs; (a) G2-20, (b) G2-15, (c) G2-10, and (d) G2-5.

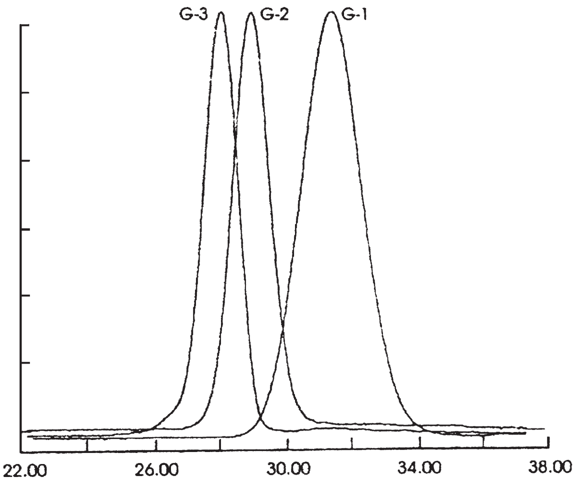


Figure 2
SEC chromatograms of three generations of dendrimer-like star polymers.

have the same magnitudes as that observed for the first and second generations, even though the number average MWs by ^{1}H-NMR show the expected values. Small angle neutron scattering (SANS) measurements for the first and second generation polymers revealed that the radius of gyration scaled with the arm functionality (f) as $f^{2/3}$ in accordance with the Daoud–Cotton model for a multiarm star polymer (Trollsås *et al.* 2000). Conversely, the hydrodynamic

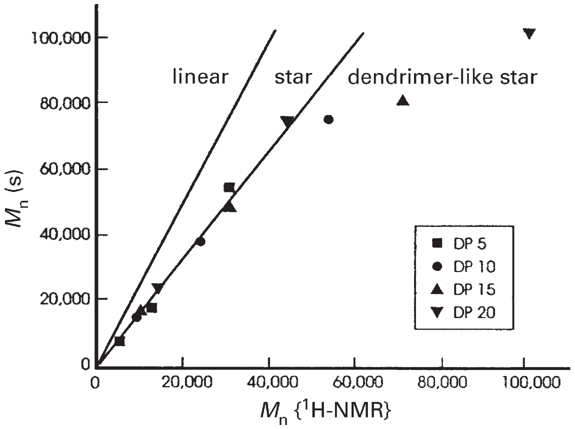
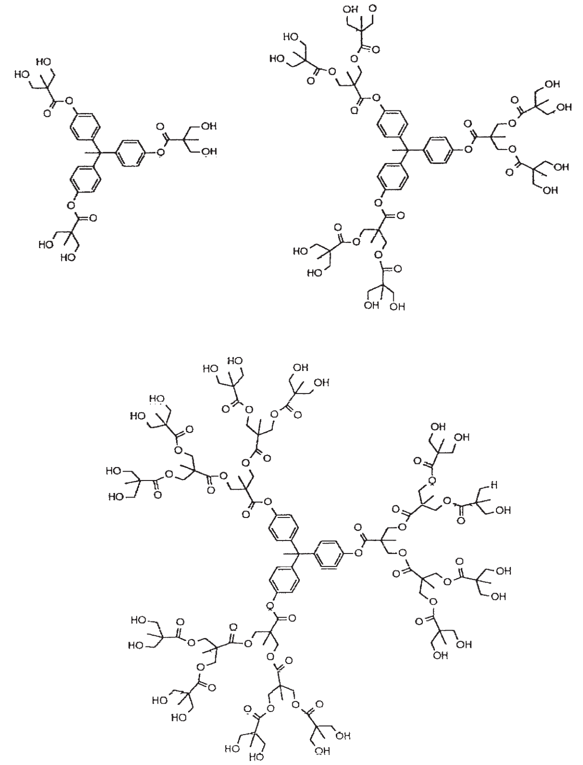


Figure 3
Number average MW from SEC vs. number average MW from NMR.

volumes and radius of gyration of the dendrimer-like star polymers with three generations of poly(ε-caprolactone) were considerably lower than their star or linear analogs. The number average molecular weights or the hydrodynamic volumes (SEC) for three generations of the dendimer-like star polymers are plotted as a function of the experimental MW (^{1}H-NMR) in Fig. 3. A gain all polymers show hydrodynamic volumes lower than linear PCL of equal MW. For polymers of different generations but with uniform DP, it was found that the deviation from linearity for the third generation polymers was more pronounced than the polymers from the first and second generation poly(ε-caprloactone). This nonlinear trend has been observed for dendrimers; however, in dendrimers the deviation from linearity has been reported to occur between the fourth and fifth generations. Polymerization appears to amplify the nonlinear characteristics observed in dendritic polymers. Clearly, the conformation of the dendrimer-like star architecture cannot be described by a traditional star model, and conformations which are considerably more space-filled than a star shaped macromolecule are produced.

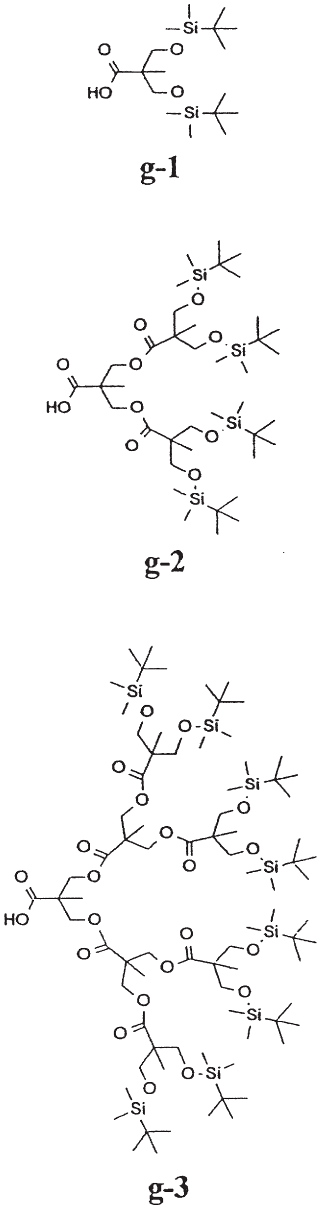
3. *Properties*

The poly(ε-caprolactone) and poly(L-lactide) dendrimer-like star polymers were semicrystalline and manifested melting temperatures and enthalpies similar to the linear and star-shaped macromolecules of comparable MWs (Trollsås and Hedrick 1998, Atthoff *et al.* 1998). Likewise, each of the polymers showed T_g values comparable to their linear analogs. Unlike their dendrimer counterparts, the higher MW dendrimer-like star polymers were film-forming and fingernail creasable, characteristic of tough ductile mechanical properties. Films of the polymers derived from poly(ε-caprolactone) and poly(L-lactide) were opaque, due to their semicrystalline morphology, whereas films prepared from the amorphous poly (*rac*-lactide) were clear.



Scheme 4

Other "building blocks" for the construction of dendrimer-like star polymers included hydroxy functionalized bisMPA dendrimers (Scheme 4), used as "initiators" and prepared by procedures developed by Hult and coworkers (Ihre *et al.* 1998) producing 6 to 48 arm star polymers, respectively, and up to three generations of tert-butyldimethylsilyl protected dendrons of bisMPA (Scheme 5) (Trollsås *et al.* 1997). In this way, macromolecular architectures that closely resemble the most progressive dendrimers may be obtained. For example, layered dendrimer-like star polymers have been prepared by modifying six arm star poly(ε-caprolactone) with various generations of amorphous *bis*MPA dendrons as the second generation, followed by the deprotection of the *tert*-butyldimethylsilyl groups and growth of additional poly(ε-caprolactone). Layered dendrimer-like star polymers with up to 48 arms and with MWs as high as 250 000 g mol^{-1} were prepared with polydispersities in the 1.10 range (Scheme 6) (Trollsås *et al.* 1998).



Scheme 5

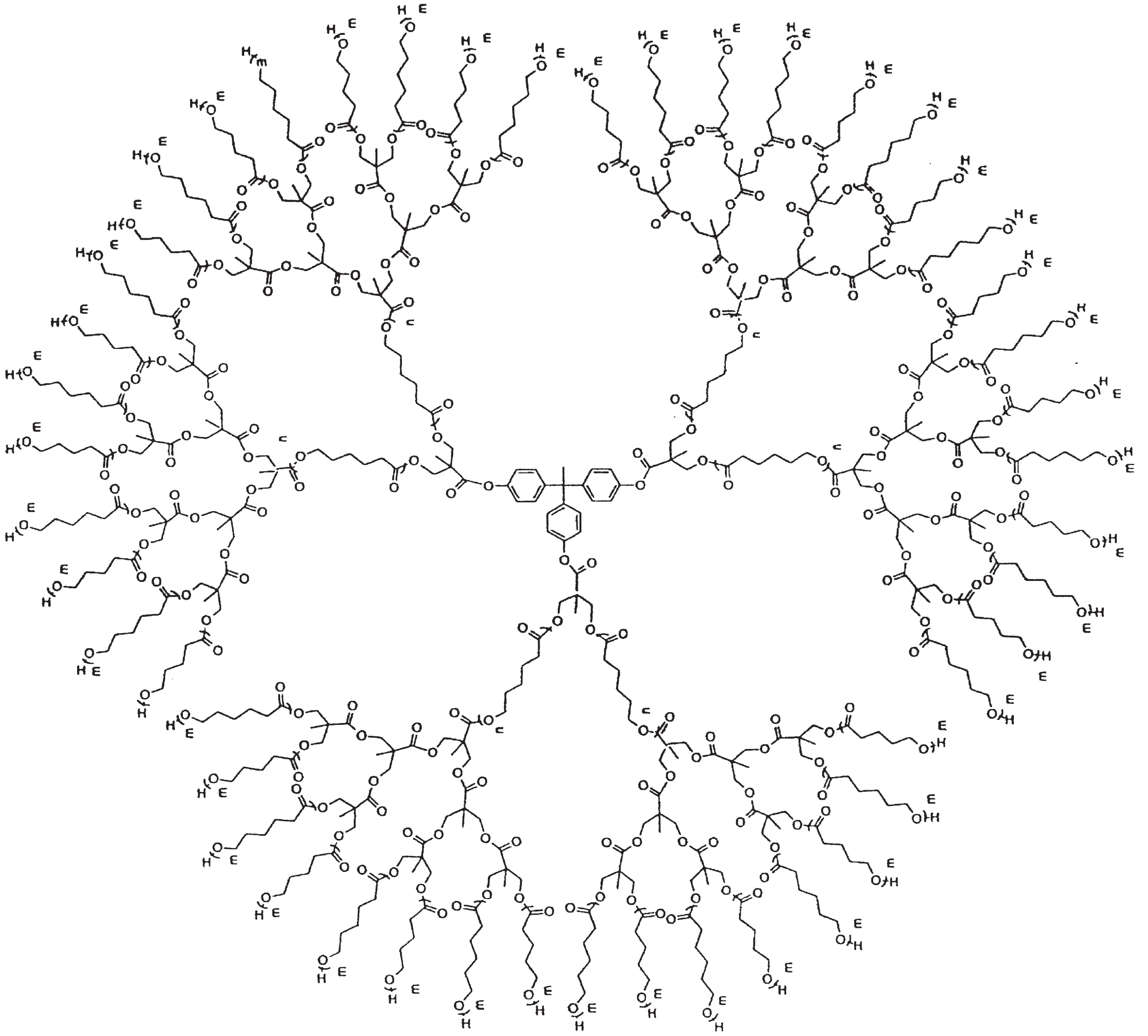
These polymers demonstrated a unique combination of properties: abundant chain end functionality and globular conformation in solution, characteristic of a dendrimer, yet with mechanical properties comparable to linear poly(ε-caprolactone).

The versatility of this modular approach to polymer design and synthesis, provided the opportunity to systematically address some of the issues of branching in macromolecules. Six constitutional or, more appropriately, topological isomers of dendrimer-like star polymers derived from derivatives of *bis*MPA and poly(ε-caprolactone) were prepared having a number average MWs of $\sim 80\,000\,\mathrm{g\,mol^{-1}}$ with exactly 45 branching junctures and 48 hydroxyl end groups (Trollsås *et al.* 2000). The only difference in the polymers is the placement in the branching junctures. The six isomers are shown in Scheme 7. The design of the six isomers is generically shown in Scheme 8, where the polymers are arranged in a triangle. From the top to the bottom left corner of the triangle, each of the polymers have six arm star poly(ε-caprolactone) as the central core, initiated from the first generation dendrimer, denoted as line 6-OH. Likewise, the line denoted as 12-OH represents the polymers initiated from the second generation dendrimer, whereas the line 24-OH denotes the polymer initiated from the third generation dendrimer.

Moving down the triangle from the top to the bottom right corner, these polymers have the first generation dendron at the polymer surface (line g-1) which is required to bring the total surface functionality to 48 and the total branching junctures (bisMPA units) to 45. Likewise, lines denoted as g-2 and g-3 are the star polymers that have either the second or third generation dendrons at the polymer surface, respectively. The horizontal lines denoted as G-1 through G-3 indicate the number of generations of poly(ε-caprolactone) or the degree of internal branching. The MWs of the isomers were in the proximity of the targeted $80\,000\,\mathrm{g\,mol^{-1}}$ (within $1500\,\mathrm{g\,mol^{-1}}$) with narrow polydispersities (1.08–1.19). SANS measurements yielded valued of the radius of gyration, R_g, which, when combined with the hydrodynamic volume measurements obtained by SEC, confirmed that the radial distribution of molecular material was consistent with the designed architectures. The most pronounced effects on the solution properties was for those polymers in which the branching was distributed throughout the polymer in a dendrimer-like fashion. Likewise, the melting points and degree of crystallinity was strongly dependent on the number of generations of poly(ε-caprolactone).

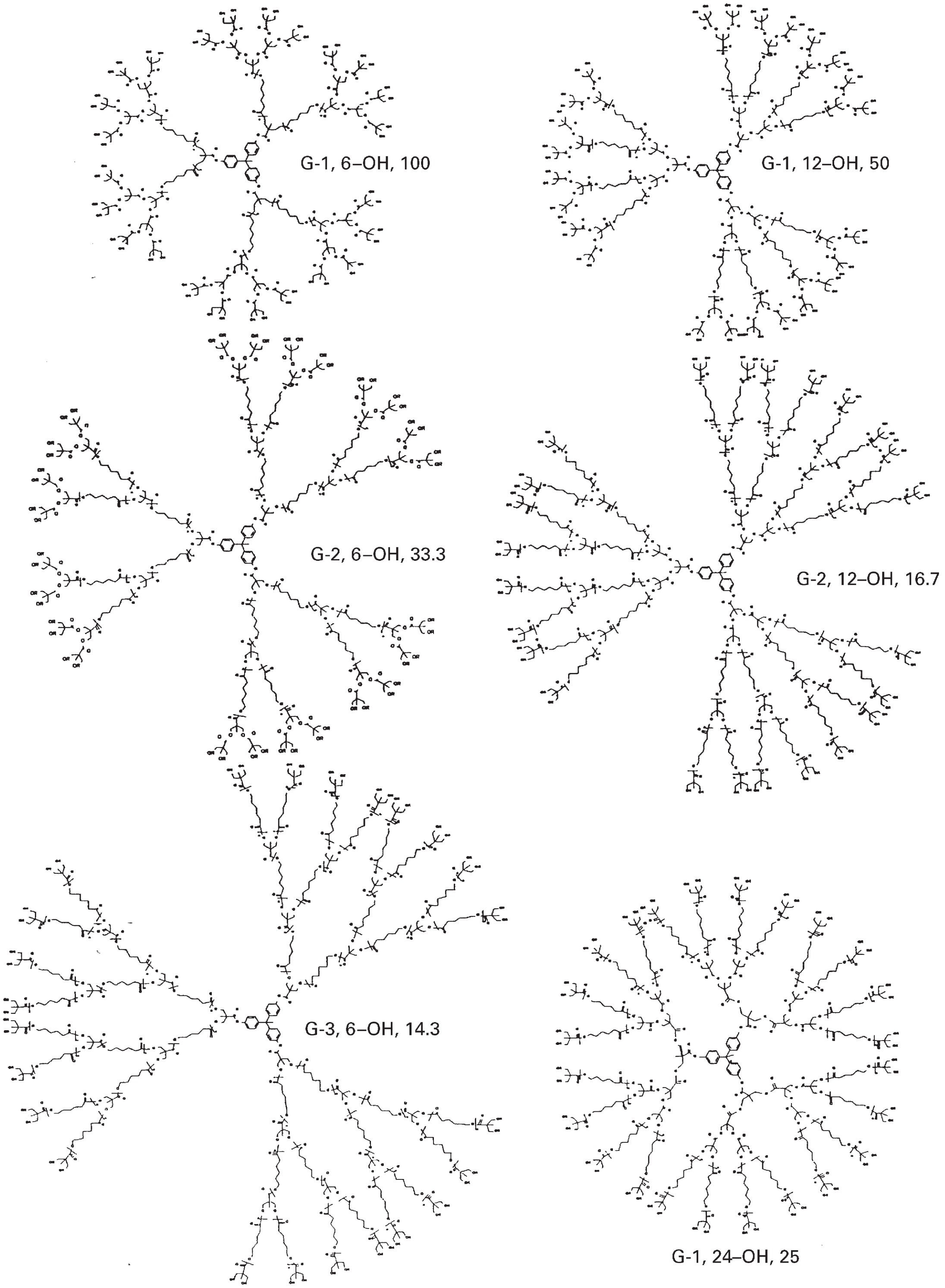
4. *Block Copolymers*

Another means of nanoscopically tailoring the dendrimer-like star polymers is by the preparation of block copolymers. The earliest dendrimer-like star block copolymers were reported by Gnanou and coworkers and were based on anionically polymerized poly(styrene) and poly(ethylene oxide). Due to the limited availability of monomers which lead to aliphatic polyesters with substantially different thermal, solution or mechanical properties, new monomers have been designed and synthesized to allow the preparation of block copolymers. For example, as a means of introducing morphology variations and desirable mechanical properties to dendrimer-like star polymers,

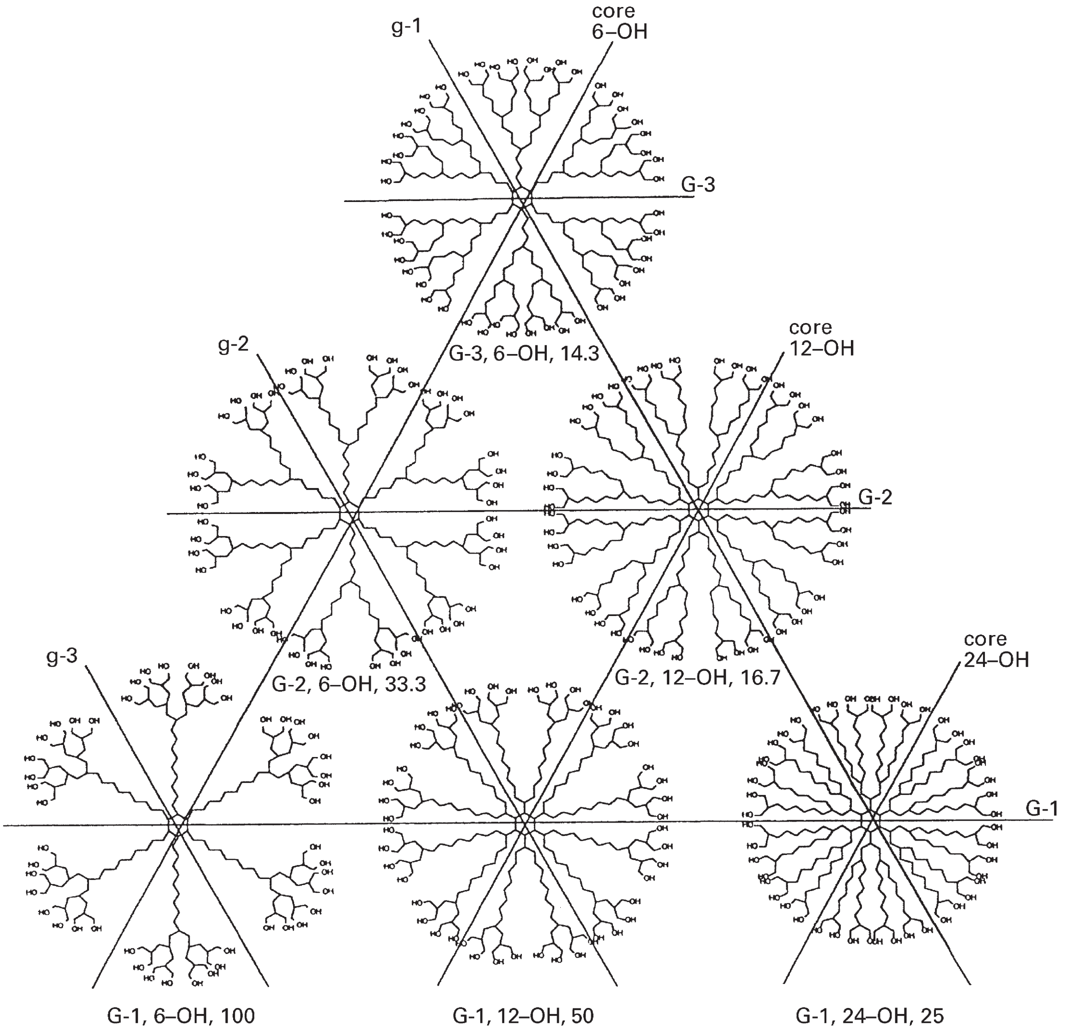


Scheme 6

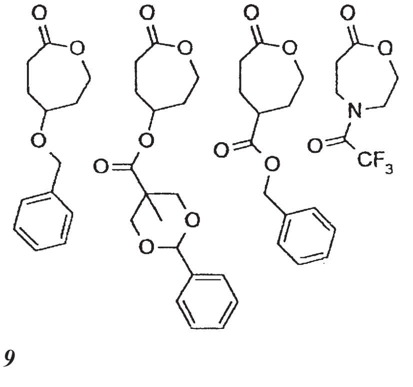
monomers bearing various substituents were prepared to preclude crystallization. This allows dendrimer-like star polymers with properties ranging from thermoplastic elastomers to rubber toughened polyesters, depending on the relative composition of the monomers to be prepared (Trollsås *et al.* 1999). Monomers containing methyl, dimethyl, ethyl, or phenyl substituents were prepared by the oxidation of the corresponding cyclohexanol or cyclohexanone (Scheme 9). Polyesters prepared from these monomers are low-T_g materials with no evidence of crystallinity. For optimum mechanical properties, the dendrimer-like star polymers were prepared with the interior generations of the copolymer comprised of the low-T_g, amorphous component, and the outer generations comprised of the high-T_g or semicrystalline component. For example, the ethyl substituted ε-caprolactone was used as the monomer for the first generation, initiated from the first generation dendrimer of bisMPA which produced a six arm polymer. Functionalization with benzylidene protected *bis*MPA, followed by deprotection, produced a six arm star polymer with 12 hydroxyl groups. This star polymer was used to initiate 12 arms of poly(L-lactide) or poly(ε-caprolactone). Microphase separated morphologies were observed for those samples derived from poly(L-lactide), as two T_g were observed by dynamic mechanical and calorimetry measurements. However, the transitions were broad and diffuse, characteristic of diffuse phase boundaries. This is not


G-1, 6–OH, 100
G-1, 12–OH, 50
G-2, 6–OH, 33.3
G-2, 12–OH, 16.7
G-3, 6–OH, 14.3
G-1, 24–OH, 25

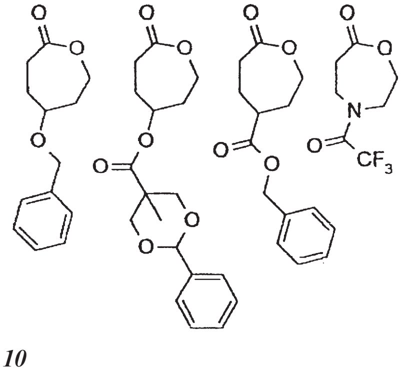
Scheme 7


g-1
core
6–OH
G-3
G-3, 6–OH, 14.3
g-2
core
12–OH
G-2
G-2, 6–OH, 33.3
G-2, 12–OH, 16.7
g-3
core
24–OH
G-1
G-1, 6–OH, 100
G-1, 12–OH, 50
G-1, 24–OH, 25

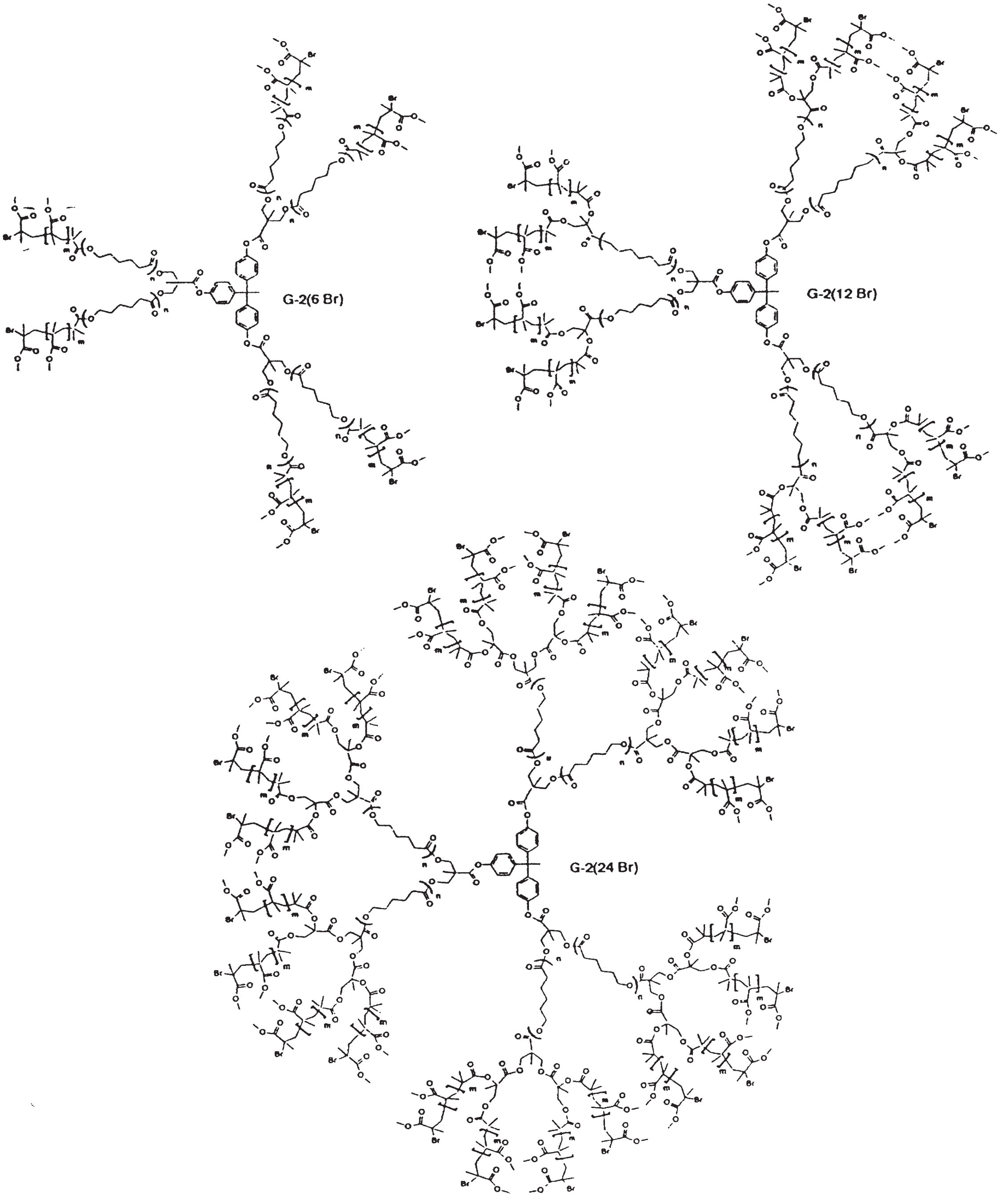
Scheme 8



Scheme 9



Scheme 10



Scheme 11

surprising considering the block lengths were between 2000 and 3000 g mol^{-1}. The mechanical properties could be controlled simply by designing either the inner or outer component to be the continuous phase by adjusting the relative degrees of polymerization of each block. Other substituted lactones developed include monomers with protected functionality (hydroxyl-, *bis*(hydroxyl)-, amino- and carboxyl) for the preparation of amphiphilic dendrimer-like star polymers (Scheme 10) (Trollsås 2000).

Structural variation in the dendrimer-like star block copolymers was also accomplished by combining ROP with other living/controlled polymerization methods such as atom transfer radical polymerization (ATRP) (Hedrick *et al.* 1998). The large number of acrylate and methacrylate monomers which can be polymerized via ATRP allowed significant synthetic flexibility in tuning the properties of the individual layers of the dendrimer-like star polymers. Capping the chain ends of the poly(ε-caprolactone) star polymers with activated bromide moieties produced the requisite macroinitiators useful for subsequent ATRP. High MW polymer was achieved with surprisingly narrow polydispersities (<1.20) (Scheme 11). Hydrolytic cleavage of the dendrimer-like star polymers released the poly(methyl methacrylate)-based tethered arms which were found to be close to their targeted MW and manifested extremely low polydispersities. Microphase separated morphologies were observed when high MW methyl methacrylate block lengths were employed. In some cases, these dendrimer-like star polymers were found to respond to changes in the polarity of the solvent (^{1}H-NMR) and serve as micelle mimics.

5. Concluding Remarks

Dendrimer-like star polymers are a new macromolecular architecture characterized by a radial geometry where the different layers or generations are comprised of high MW polymer emanating from a central core. Unlike traditional dendrimer synthesis where the molecular weight build up is slow and tedious, the use of monodispersed macromolecular building blocks allows a rapid increase in MW in just a few generations. A key advantage to the use of macromolecular building blocks is the simple purification required between transformations (i.e., polymer precipitation). Considerable versatility in the morphology of these novel homo- and block polymers is realized by variation of monomer type, block lengths, etc. It is anticipated that future work in this rapidly expanding field of hybrid dendritic-linear block copolymers will further demonstrate the unique physical properties of these novel macromolecular architectures.

Bibliography

Angot S, Murthy K S, Taton D, Gnaou Y 1998 Atom-transfer radical polymerization of styrene using a novel octafunctional initiator. Synthesis of well defined polystyrene stars. *Macromol.* **31**, 7218–25

Atthoff B, Trollsås M, Claesson H, Hedrick J L 1998 Poly(lactides) with controlled molecular architecture initiated from hydroxyl function dendrimers and the effect on the hydrodynamic volume. *Macromol. Chem. Phys.* **200**, 1333–9

Fréchet J M J 1994 Functional polymers and dendrimers. *Science* **263**, 1710

Frey H, Lach C, Lorentz K 1998 Heteroatom-based dendrimers. *Adv. Mater.* **10**, 279–93

Hawker C J, Malmström E E, Frank C W, Kampf J P 1997 Exact linear analogues of dendritic polyether macromolecules. Design, synthesis and unique properties. *J. Am. Chem. Soc.* **119**, 9903–4

Hecht S, Ihre H, Fréchet J M J 1999 Porphyrin core star polymers: synthesis, modification and implication for site isolation. *J. Am. Chem. Soc.* **121**, 9239–40

Hedrick J L, Trollsås M, Hawker C J, Atthoff B, Claesson H, Heise A, Miller R D, Jérôme R, Dubois P 1998 Dendrimer-like star block and amphiphilic copolymers by combination of ring-opening and atom transfer radical polymerization. *Macromol.* **31**, 8691

Ihre H, Hult A, Gitsov I, Frechet J M J 1998 Double stage convergent approach for the synthesis of functionalized dendritic aliphatic polyesters based on 2,2-bis(hydroxymethyl)propionic acid. *Macromol.* **31**, 4061–8

James D E 1996 In: *Encyclopedia of Polymer Science*. Wiley, New York, Vol. 6, p. 383

Johnson L K, Killian C M, Brookhart M 1995 New Pd(11)-based and Ni(11)-based catalysts for polymerization of ethylene and α-olefins. *J. Am. Chem. Soc.* **117**, 6414–5

Mio C, Kiritsov S, Thio Y, Brafman R, Prausnitz J, Hawker C J, Malmström E 1998 Vapour-liquid equilibria for solutions of dendritic polymers. *J. Chem. Eng. Data* **43**, 541–50

Percec V, Ahn C-H, Ungar G, Yeardley D J P, Moller M, Sheiko S S 1998 Controlling polymer shape through the self-assembly of dendritic side groups. *Nature* **391**, 161–4

Roovers J, Zhou L L, Toporowski P M, Van der Zwan M, Iatrou H, Hadjichristidis N 1998 Regular star polymers with 64 and 128 arms: models for polymeric micelles. *Macromol.* **31**, 6762

Six J L, Gnanou Y 1995 From star-shaped to dendritic poly(ethylene oxide)s: towards increasingly branched architectures by anionic polymerization. *Macromol. Symp.* **95**, 137–50

Trollsås M, Atthoff B, Würsch A, Hedrick J L, Pople J A, Gast A P 2000 Constitutional isomers of dendrimer-like star polymers: Design, synthesis, and confirmational and structural peoperties. *Macromolecules* **33**, 6423

Trollsås M, Claesson H, Atthoff B, Hedrick J L 1998 Layered dendrimer-like star polymers. *Angew Chem., Int. Ed. Engl.* **37**, 3122

Trollsås M, Claesson H, Atthoff B, Hedrick J L, Pople J A, Gast A P 2000 Confirmation and structural properties of highly functional dendrimer-like star polymers synthesized from living polymerization techniques. *Macromol. Symp.* **153**, 87

Trollsås M, Claesson H, Hedrick J L, Pople J A, Gast H P 2000 Dendrimer-like star polymers: amplification of star polymer properties in structure and confirmation. *Macromol.* in press

Trollsås M, Hedrick J L 1998 Dendrimer-like star polymers. *J. Am. Chem. Soc.* **120**, 4644

Trollsås M, Hedrick J L, Mecerreyes D, Dubois P H, Jérôme R, Ihre H, Hult A 1997 Versatile and controlled synthesis of star and branched macromolecules by dendritic initiation. *Macromol.* **30**, 8508–11

Trollsås M, Lee V Y, Mercerreges D, Löwenhielm P, Müller M, Miller R D, Hedrick J L 2000 Hydrophilic aliphatic polyesters: Design, synthesis and ring-opening polymerization. *Macromolecules* **33**, 4619

Trollsås M, Kelly M A, Claesson H, Siemens R, Hedrick J L 1999 Highly branched block copolymers: design, synthesis and morphology. *Macromol*. **32**, 4917–24

Ueda J, Kamigaito M, Sawamoto M 1998 Calixarene-core multifunctional initiators for the ruthenium living radical polymerization of methacrylates. *Macromol*. **31**, 6762–8

Webster O W 1994 Living polymerization methods. *Science* **251**, 887

J. L. Hedrick
IBM Almaden Research Center, San Jose, California, USA

M. Trollsås
Candescent Technologies, San Jose, California, USA

This Page Intentionally Left Blank

I

Ice: Structures

There are at least twelve known phases of ice (Fig. 1), with several others suspected to exist. All the known phases consist of 4-coordinated networks of water molecules linked through hydrogen-bonded interactions. In addition to the underlying oxygen frameworks, understanding these structures requires a knowledge of the way in which the hydrogen atoms of the water molecules "decorate" this framework. These arrangements may be disordered, ordered, or partially ordered, and the degree of order may change with pressure or temperature.

The structures of most of the known phases are reasonably well understood. A range of topologies is seen, some of which have counterparts in other four-coordinated systems such as silicates, together with wide variations in intermolecular bond angles and distances that are beginning to help us understand some of the forces that control hydrogen ordering. At very high pressures, theory predicts that the material loses its molecular nature; experimental work supports these predictions. At ambient pressure, questions remain about the hydrogen disorder in the common phase ice Ih. Much of what is known about the structures and properties of ice is covered in a monograph by Petrenko and Whitworth (1999).

1. *The Water Intermolecular Potential*

Any structure must be consistent with the potential function describing the intermolecular interactions. The structure adopted will depend on how these interactions respond to external temperature and pressure constraints. Although the details of the water–water hydrogen-bonding interaction remain controversial, an acceptable operational picture emphasizes a tetrahedral interaction geometry, with two protons able to donate to the lone-pair region of each of two neighboring molecules. In addition to the directional attraction inherent in the various hydrogen-bonding models, the repulsive core of the potential is important, especially when considering structures under pressure. Although a spherically symmetrical

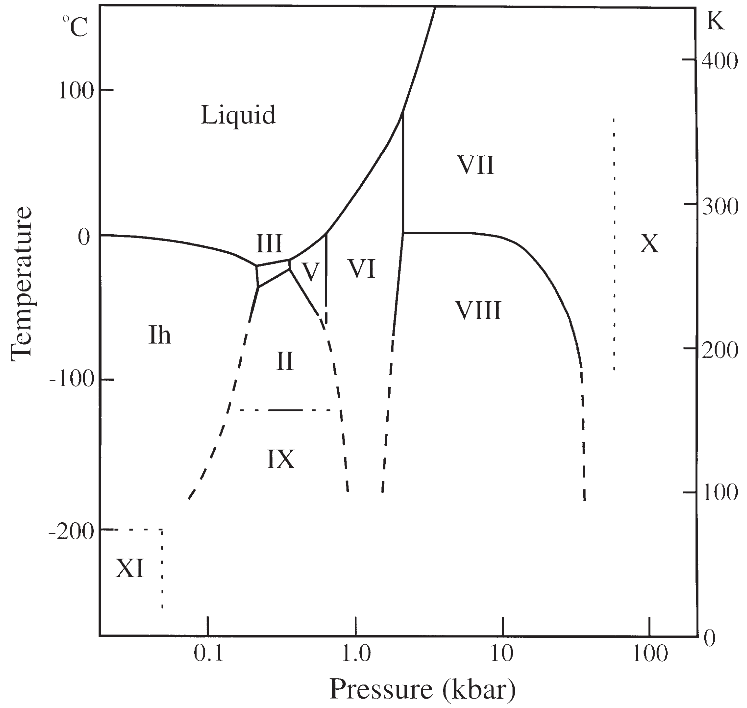


Figure 1
The phase diagram of ice showing the known stable phases. Note the logarithmic pressure scale. Solid lines represent measured transitions; thick dashed lines represent transitions extrapolated to low temperatures; thin dashed lines represent predicted transitions.

repulsion is usually assumed, there is evidence that this is an oversimplification (Savage 1986, Savage and Finney 1986); a detailed understanding of ice and other aqueous structures will need to consider a small, but significant, anisotropic repulsion.

2. *Structures: Frameworks*

It is instructive to consider the ice structures initially in terms of the oxygen frameworks. Links between the oxygens denote hydrogen-bonding interactions, with one hydrogen located between each pair of connected oxygens. As the structures deform under pressure, the hydrogens move further away from the O···O line, but the degree of distortion does not cause ambiguity in framework identification.

2.1 Ambient Pressure

The phase diagram of Figure 1 indicates the existence of one ambient-pressure phase, the common ice Ih of the freezer, glaciers, and snowflakes. Its framework structure (Fig. 2) is essentially that of the hexagonal wurtzite structure, consisting of an ABABAB stacking of layers of (chair) puckered hexagonal rings. The ideal tetrahedral angle of 109.5° is close to the 104.5° H–O–H angle in the water molecule, resulting in a relatively unstrained structure. The open structure, a consequence of the tetrahedral intermolecular geometry, should be noted. A detailed description of the ice Ih structure is given by Kuhs and Lehmann (1986).

A cubic variant of this structure, commonly termed *cubic ice*, or ice Ic, can be prepared by a range of methods, including from higher pressure phases. Although this structure is generally considered to be a separate phase, consisting of an ABCABC stacking of the same puckered hexagonal ring planes as found in ice Ih (similar to the zinc blende structure), no clean diffraction pattern has been reported to fully verify this. Rather, diffraction data indicate considerable disorder in the structure, together with small crystallite size (Kuhs *et al.* 1987). This structure may possibly be better described as a highly stacking-faulted ice Ih. It may also be the preferred structure only in small crystallites. As crystal growth proceeds, the ABCABC structure may become unstable with respect to the ABABAB stacking, with a subsequent transformation to Ic. This interpretation is consistent with several experimental results, for example the observation of cubic ice in confined volumes (e.g., in porous vycor with pore sizes of tens of angstroms), and the annealing of the disorder between about 160 and 210 K, followed by the formation of ice Ih above this temperature.

2.2 Higher Pressure Frameworks

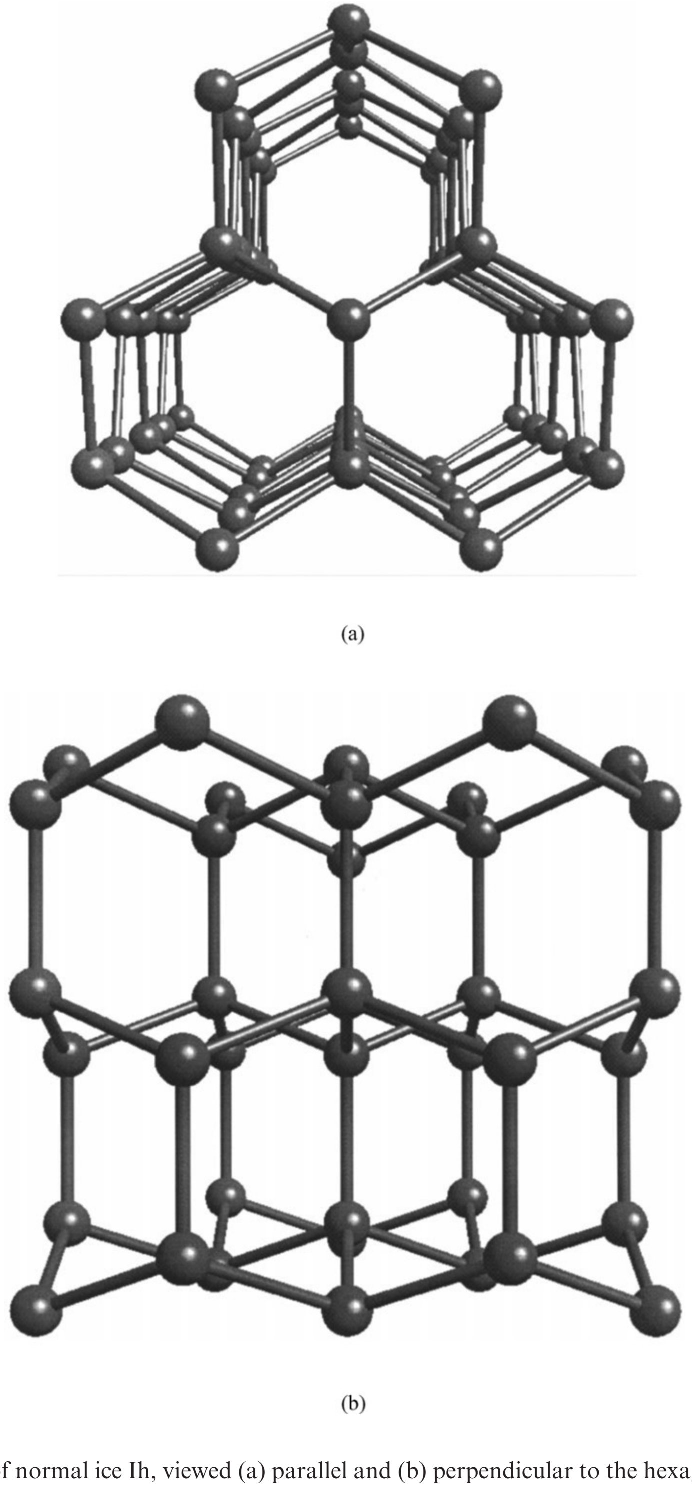
As ice Ih is subjected to pressure, the molecules must rearrange to occupy less volume, but in a way that is still consistent with the intermolecular interactions. The high-pressure structures might thus be rationalized in terms of the different (still four-coordinated) networks that can be formed with an increased degree of O–O–O bond bending. The oxygen topologies of the stable ice phases which are found below about 0.6 GPa, namely ices II, III, V, VI, VII, VIII, and IX, are now established (Petrenko and Whitworth 1999, Lobban 1998). They contain a range of ring structures, including highly strained four-fold rings in, for example, ice V. Phases II, III, V, and IX (which has an oxygen topology identical to that of ice III) accommodate the increased pressure through forming networks that show increased O–O–O distortion. On increasing the pressure further, the water molecules solve their internal rearrangement problem by taking advantage of the inherently low densities of four-coordinated tetrahedral networks. This "excess free volume" is taken up by forming denser structures consisting of two interpenetrating but (hydrogen-bond) disconnected sublattices, with consequent reduced O–O–O bond bending. For example, ices VII and VIII (which again have the same oxygen topology) can be thought of as two interpenetrating diamond-type lattices (Fig. 3).

Interestingly, in the dense ice VIII structure, the shortest nonhydrogen-bonded O···O distance between neighbors in different networks is less than the hydrogen-bonded distance within each network (Kuhs *et al.* 1984). Although initially a surprising observation, this is entirely consistent with the anisotropy of the water molecule repulsive core (Savage and Finney 1986). Savage (1986) has commented on the importance of repulsions in determining the pressure stability limits of ice structures, relating the pressure-induced transitions to relieving the strain induced in certain specific repulsive interactions.

2.3 Metastable Phases

Bridgman's (1935) pioneering measurements of the ice phase diagram identified a metastable phase in the region of stability of ice V. This he called ice IV. More recently, Lobban *et al.* (1998) found and solved the structure of a second presumed metastable phase (ice XII) in the same region of the phase diagram. There is evidence that at least one, and possibly more, further phases may exist in or around this region (Lobban 1998, Mishima and Stanley 1998). Thus, in solving its problem of how to arrange its molecules to satisfy its own intermolecular interactions and the external constraints, the system finds at least three "similarly satisfactory" structural solutions in this pressure/temperature region.

It is interesting to note some of the elements of these three "structural solutions." First, the densities of the metastable phases IV and XII are at 1.4361(1) g cm^{-3} and 1.4365 (1) g cm^{-3} respectively, both (i) almost the same and (ii) *greater* than the



(a)

(b)

Figure 2
The oxygen framework of normal ice Ih, viewed (a) parallel and (b) perpendicular to the hexagonal channels. Note the open structure.

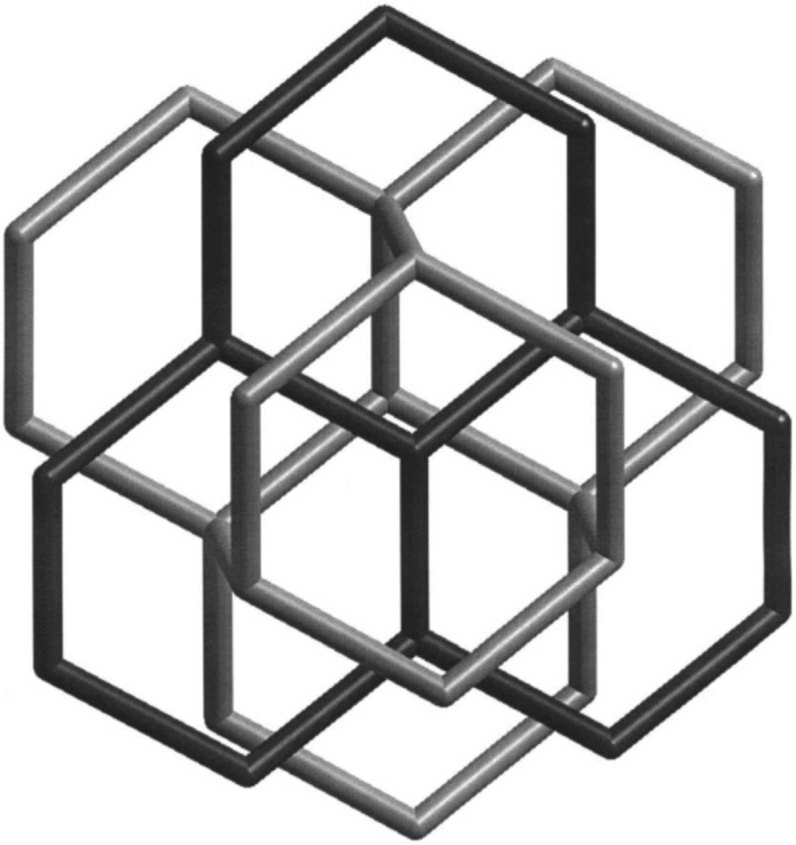


Figure 3
The ice VII structure in which two diamond-type structures interpenetrate each other. The two sublattices are shaded differently for clarity.

density of the stable phase ice V (1.4021(1) g cm^{-3}). Furthermore, if we estimate the strain in these structures through the average (root mean square) degree of O–O–O bond bending, $\delta\theta$, from the ideal tetrahedral angle, we find that ice IV—a metastable phase—shows the least average distortion ($\delta\theta = 15.06^\circ$), while for the stable ice V, $\delta\theta$ is higher at 18.54° (similar to the 18.87° of the other metastable phase XII). This apparently odd trend in bond-angle distortion can be at least partly understood by noting that the single network structure of ice IV shows a threading of one of the O–O hydrogen bonds through a hexagonal ring of other hydrogen-bonded molecules (Fig. 4). The system seems to use this "partial self-clathrate" arrangement to relieve some of the distorted oxygen bond-angle strain.

The fineness of the free energy balance in these structures is underlined by the small free energy differences between the three phases (Lobban 1998), which are estimated to be of the order of 50–100 J mol^{-1}. The reliable calculation of such small free energy differences from first principle calculations is presently beyond our capabilities. We are therefore observing some really quite subtle effects which have significant structural consequences. In this case, as in some other ice phases, the ordering or otherwise of the hydrogen atoms on the oxygen framework, and the driving forces underlying these orderings (see Sect. 3), are likely to significantly influence structural stability.

2.4 *Very High Pressure Phases*

Ices VII and VIII (Fig. 3) are relatively dense body-centered oxygen structures. Increasing pressure further is therefore unlikely to result in network reconstruction. Rather, the O–O distance might be shortened to the point at which the hydrogen is no longer preferentially attached to one of the two oxygens forming the relevant hydrogen-bonded link. We thus expect to approach a nonmolecular structure in which the hydrogen is found halfway between the two oxygens, which will be considerably closer than the normal hydrogen-bonded distance. This structure has been named ice X. Spectroscopic (Goncharov

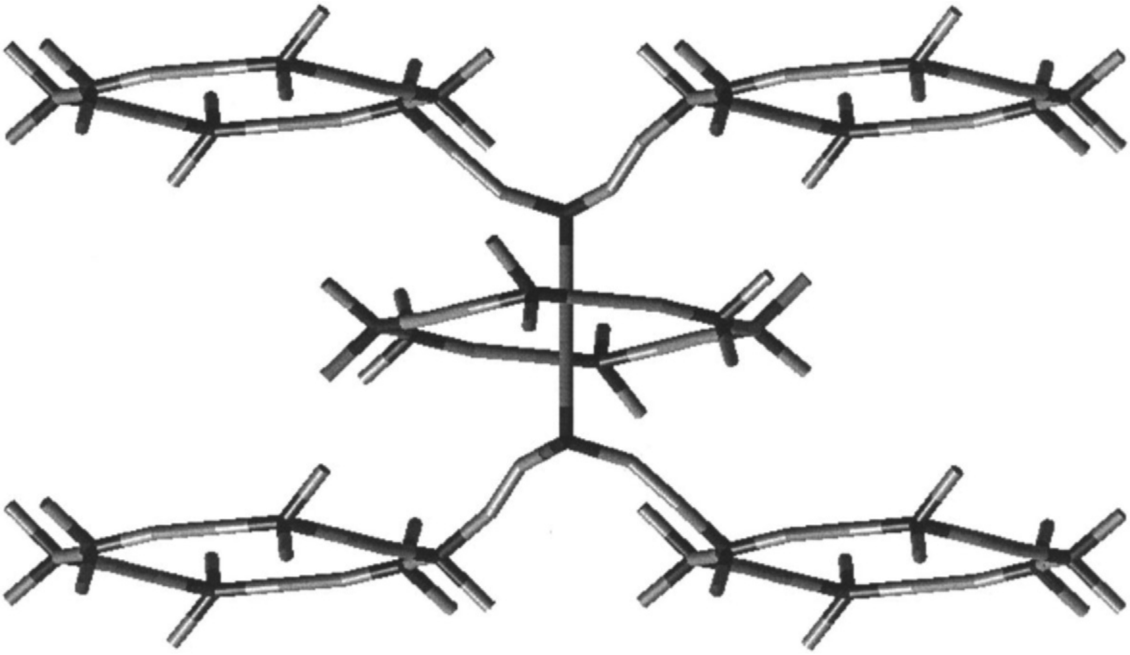


Figure 4
Part of the structure of ice IV, showing a hydrogen bond passing through the center of a hexagonal ring.

et al. 1996) and direct x-ray (Loubeyre *et al.* 1999) experimental evidence that this does occur at pressures of around 60 GPa has been obtained. In addition, *ab initio* quantum mechanical calculations predict such a transition at around 70 GPa (Benoit *et al.* 1998). The approach to this transition appears to be quite complex, with the suggestion of a sequence of "transitional" spatially modulated phases relating to the ordering of ice VII occurring between 2.2 and 25 GPa. There is also evidence that a further change may be taking place above 150 GPa. It is speculated that this could be a displacive transition along a predicted body-centered cubic to hexagonal close-packed path for the oxygens (Benoit *et al.* 1996).

3. Structures: Hydrogen Positions

3.1 Orientational Order and Disorder

To complete the structural picture, the oxygen frameworks discussed above need to be supplemented by a knowledge of how the hydrogen atoms are arranged on these frameworks.

The hydrogen decoration of the oxygen frameworks must fulfil the topological constraints of the Bernal–Fowler rules. These state that (i) each oxygen is associated with two hydrogen atoms, consistent with the existence of individual water molecules, (ii) each O–O hydrogen bond accommodates one, and only one, hydrogen, and (iii) each water molecule links to four neighbors through hydrogen bonding. Within these constraints, there are (in general) a number of ways in which the hydrogens can be distributed on a four-coordinated oxygen framework or, to make an equivalent statement, there are a number of ways in which the water molecules can be oriented. These arrangements may be ordered (e.g., ferroelectric, antiferroelectric), or disordered (e.g., where there is no periodicity of the orientations that coincides with the periodicity of the oxygen framework). Subject to crystallographic symmetry constraints, there are also possibilities of intermediate partially ordered states.

This is illustrated in Fig. 5(a), which shows the six possible orientations that a water molecule can adopt such that it donates two and accepts two hydrogen bonds to and from its neighbors. Figure 5(b) shows schematically in two dimensions both an orientationally ordered (here ferroelectrically ordered) and an orientationally disordered structure. Note that the space–time average structure, as determined by diffraction, will be the average of all the orientations in a given structure. Thus, a fully disordered structure will appear as a tetrahedral network of oxygen atoms with *two* hydrogen atoms per O–O bond; the probability of finding a hydrogen at each site is, however, only 50% (Fig. 5(c)). For a partially ordered structure, these probabilities will depart from 50%.

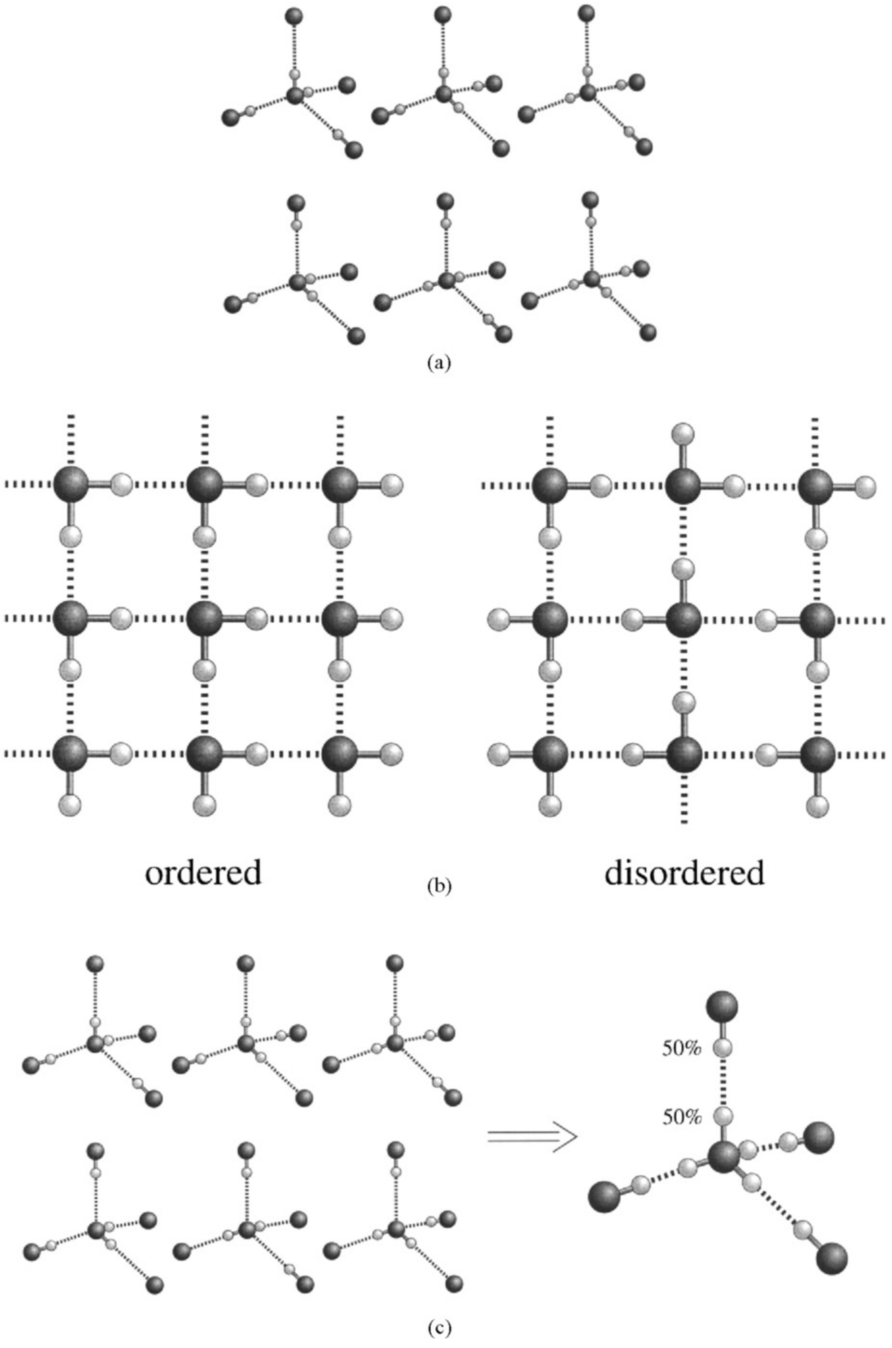


Figure 5
(a) Possible orientations of a water molecule hydrogen bonded to four neighboring water molecules; (b) an example schematic ice structure with the water molecules ordered (left) and disordered (right); (c) the space–time average structure of a fully disordered structure as seen by diffraction.

Early evidence for hydrogen ordering in ice structures came from physical measurements such as dielectric and thermodynamic properties. This argued that all the then-known ice phases were hydrogen disordered except II, IX (thought to be a low-temperature hydrogen-ordered form of ice III), and VIII. Neutron studies, which are sensitive to the hydrogen positions, have since confirmed the ordered states of ices II (Londono *et al.* 1992, Lobban 1998) and VIII (Kuhs *et al.* 1984), though they have also questioned whether ices III (Londono *et al.* 1993) and V (Hamilton *et al.* 1969) are fully disordered. Further work (Lobban *et al.* 1998, 2000, Kuhs *et al.* 1998) has confirmed that, contrary to being fully disordered as expected, both III and V are partially ordered even at temperatures close to their melting lines. In the case of ice V, the degree of ordering is pressure and temperature dependent. There is also evidence that ice IX is not fully ordered (Londono *et al.* 1993), though at the low temperatures involved, the residual order may be kinetically frozen in (see Sect. 3.4 below).

3.2 Oxygen and Hydrogen Disorder

Although disorder in ices is normally considered only with respect to the hydrogen positions, there can be consequential disordering effects on the oxygen positions. These were first pointed out in the case of ice VII, where the neutron data could be explained by the oxygen of the water molecule being shifted a small amount from the crystallographic mean position to accommodate the local strains caused by the water molecule orientation (Kuhs *et al.* 1984). The magnitude of this displacement is about the same as the distance between the oxygen atom and the centre of mass of the molecule. Similar oxygen disorder was subsequently found in ice VI (Kuhs *et al.* 1989) and even in the ambient phase Ih (Kuhs and Lehmann 1986).

3.3 Forces Controlling Orientational Order and Disorder

Although there is no quantitative explanation of orientational (partial) disorder in ice structures, a plausible case for the main determining factors can be made.

Consider the local geometry of a water molecule in an ice structure, in particular the six possible O–O–O angles around this position. For this central water molecule, we can (subject to the Bernal–Fowler rules) arrange the two hydrogens in one of six ways (the number of ways of choosing two from four), with each of these choices corresponding to a water molecule orientation. To each of these possible orientations, there will be associated a certain O–O–O angle. A comparison of this O–O–O angle with the H–O–H angle of the water molecule will indicate how well these two angles match each other. Where the angular matches for all six possible orientations are similar (as in ice Ih), then there will be no energetic preference for one orientation over another. If, as is the case in the high-pressure phases, there are different degrees of angular mismatch between different possible orientations, then there is an energetic penalty for occupying those orientations with the greatest mismatch. Hence, in the simplest analysis, one might expect a preferential occupation of those orientations that show the least angular mismatch, and hence lowest strain.

In applying this rationale to real structures where the detailed geometry and degree of ordering is known experimentally, the Bernal–Fowler rules and the constraints of the crystallographic symmetry need to be considered, resulting in some quite complex arguments (Lobban 1998). However, it is clear when this is done that this angle-mismatch approach cannot be the complete story. Although the details are not understood in a single case, it seems likely that we need to consider also, in addition to the angular strain, the intermolecular repulsions which will also in general be different for different molecular orientations.

Despite the complexity of the orientational order/disorder in ice structures, it is important to gain a better understanding of the phenomenon. It is potentially a determining factor in the stability of a phase; for example, it can be plausibly argued that, were it not for the orientational disorder in ice V, ice IV would not exist.

3.4 Temperature Dependence of Orientational Order

Orientational order should increase as temperature is reduced. Therefore, ordinary ice Ih, which is fully disordered at 0 °C, should orient fully at 0 K. However, it has been realized for decades that this is not the case; Pauling's original calculation of the residual entropy confirmed quantitatively that ice Ih remains fully disordered at 0 K. To avoid contradicting the third law of thermodynamics, we must conclude that this disordered structure at low temperature is not the equilibrium structure; there must be some kinetic restriction preventing the molecules reorienting as temperature falls.

Similar considerations should apply to all the disordered ice phases. Yet, in most of these phases there is no strong evidence to support a temperature-driven ordering. A number of measurements of thermodynamic and dielectric properties have suggested in some phases—for example ices V and VI—an approach to some kind of phase transition as temperature is reduced, but structural evidence has not yet been obtained to support these suggestions. The dynamics of water molecule reorientations do become very sluggish as temperature is reduced, with

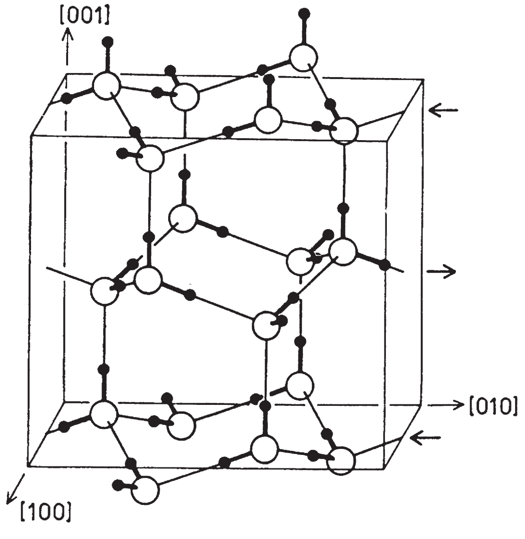


Figure 6
The hydrogen ordered structure of ice XI (orthorhombic; $Cmc2_1$).

the result that we appear to have an orientational glass in most ice structures at low temperature.

There is much active work in this area at present, but it is not possible here to discuss in any detail the factors that may control hydrogen diffusion or molecular reorientation kinetics in water. An example of relevant active work is the study of the ordering of ice Ih towards the equilibrium low-temperature ordered phase which has been named ice XI. One factor preventing water molecule reorientation is the need to make a large number of cooperative movements rather than single reorientations. This is imposed by the Bernal–Fowler constraints: in a given molecular configuration, all the O–O links are already occupied by hydrogens, so the water orientations are locked in. One way of partly unlocking this situation is to introduce defects which remove some of the hydrogens. This can be done to some degree by doping ice Ih with KOH, though its low solid-state solubility limits the degree to which these OH^- defects can be introduced. Calorimetric measurements on this doped system by Tajima *et al.* (1982) suggest the existence of a phase change around 100 k, and subsequent crystallographic work has demonstrated the existence of partial orientational ordering (Jackson *et al.* 1997). Although it has not proved possible to induce more than a small degree of such ordering, the data obtained are consistent with the low-temperature ordered structure being ferroelectric (Fig. 6).

4. *Summary*

Although ice might seem at first sight a very simple system, its phase diagram is relatively complex. When the possibilities of hydrogen disorder are also considered, the behavior of this material appears to be even more complex. Since about the mid-1980s, neutron scattering measurements have clarified the structures of most of the known phases. The orientational order/disorder situation in several phases has also been clarified, though uncertainties remain in others showing partial disorder. The factors that control this disorder may relate to both local bond angle strains and non-bonded repulsions, though quantitative predictions of observed partial disorder are not yet possible. At the highest pressures above 60 GPa or so, the evidence for a non-molecular ice phase is looking increasingly strong, though the transition to ice X appears to be not simple.

More uncertainty surrounds the kinetics of molecular rotations in all the ice phases. Below a certain temperature, perhaps around 120 K, molecular reorientations seem to essentially freeze in, preventing the disordered phases from approaching their expected low-temperature ordered structures. Active research is continuing to try to improve our understanding of both the structural and the dynamical aspects of the disorder in the various phases of ice.

Bibliography

Benoit M, Bernasconi M, Focher P, Parrinello M 1996 New high pressure phase of ice. *Phys. Rev. Lett.* **76**, 2934–6

Benoit M, Marx D, Parrinello M 1998 Tunnelling and zero-point motion in high pressure ice. *Nature* **392**, 258–61

Bridgman P W 1935 The *P–V–T* relations of the liquid, and the phase diagram of heavy water. *J. Chem. Phys.* **3**, 597–605

Goncharov A F, Struzhkin V V, Somayazulu M S, Hemley R J, Mao H K 1996 Compression of ice to 210 gigapascals: infrared evidence for a symmetric hydrogen-bonded phase. *Science* **273**, 218–20

Hamilton W C, Kamb B, La Placa S J, Prakash A 1969 In: Riehl N, Bullemer B, Engelhardt H (eds.) *Physics of Ice*. Plenum, New York, pp. 44–58

Jackson S M, Nield V M, Whitworth R W, Oguro M, Wilson C C 1997 Single-crystal neutron diffraction studies of the structure of ice XI. *J. Phys. Chem. B* **101**, 6142–5

Kuhs W F, Ahsbahs H, Londono D, Finney J L 1989 In-situ crystal growth and neutron four-circle diffractometry under high pressure. *Physica B* **156/7**, 684–7

Kuhs W F, Bliss D V, Finney J L 1987 High resolution neutron powder diffraction study of ice Ic. *J. Phys.* **48** C1, 631–6

Kuhs W F, Finney J L, Vettier C, Bliss D V 1984 Structure and hydrogen ordering in ices VI, VII, and VIII by neutron powder diffraction. *J. Chem. Phys.* **81**, 3612–23

Kuhs W F, Lehmann M S 1986 In: Franks F (ed.) *Water Science Reviews*. Cambridge University Press, Cambridge, Vol. 2, pp. 1–65

Kuhs W F, Lobban C, Finney J L 1998 Partial H-ordering in high pressure ices III and V. *Rev. High Pressure Sci. Technol.* **7**, 1141–3

Lobban C 1998 Neutron diffraction studies of ices. *PhD Thesis*. University of London

Lobban C, Finney J L, Kuhs W F 1998 The structure of a new phase of ice. *Nature* **391**, 268–70

Lobban C, Finney J L, Kuhs W F 2000 The structure and ordering of ices III and V. *J. Chem. Phys.* **112**, 7169–80

Londono D, Finney J L, Kuhs W F 1992 Formation, stability and structure of helium hydrate at high pressure. *J. Chem. Phys.* **97**, 547–52

Londono D, Kuhs W F, Finney J L 1993 Neutron diffraction studies of ices III and IX on under-pressure and recovered samples. *J. Chem. Phys.* **98**, 4878–88

Loubeyre P, Le Toullec R, Wolanin E, Hanfland M, Husermann D 1999 Modulated phases and proton centring in ice observed by X-ray diffraction up to 170 GPa. *Nature* **397**, 503–6

Mishima O, Stanley H E 1998 Decompression-induced melting of ice IV and the liquid–liquid transition in water. *Nature* **392**, 164–8

Petrenko V F, Whitworth R W 1999 *Physics of Ice*. Oxford University Press, Oxford

Savage H F J 1986 In: Franks F (ed.) *Water Science Reviews*. Cambridge University Press, Cambridge, Vol. 2, pp. 67–148

Savage H F J, Finney J L 1986 Repulsive regularities of water structure in ices and crystalline hydrates. *Nature* **322**, 717–20

Tajima Y, Matsuo T, Suga H 1982 Phase transition in KOH-doped hexagonal ice. *Nature* **299**, 810–2

J. L. Finney
University College London, UK

Injection Molded Semicrystalline Polymers: Structure Development and Modeling

A variety of molding processes is used commercially, including injection, compression, transfer, blow, and rotational molding. Which process is employed depends largely on the design, final application, and requirements of the finished product. Among these molding processes, the most important commercially is injection molding. In this process, molten polymers are injected under high pressure into a cold mold cavity in the case of thermoplastics, or into a heated mold cavity in the case of thermosets. With thermoplastics, cooling continues until the material has solidified sufficiently to permit ejection of the molded products from the mold without damage. Therefore in semicrystalline polymers the final products have complex distributions of crystalline morphology, such as skin-core structure. Skin-core structure is the typical cross-sectional morphology of an injection molded polymer. Recently, computer-aided-engineering (CAE) has become a powerful tool for investigating process analysis. CAE is useful for optimizing the molding conditions, and for helping in the design of the molds, and has several other functions such as flow analysis, packing and cooling analysis, warpage analysis, etc. These analyses of semicrystalline polymers sometimes give poor reliability, due to complex distributions of crystalline morphology. Therefore it is necessary to develop a simulation of crystallization in the injection molding process. Furthermore, this crystallization simulation may clarify the mechanism of crystallization in the injection process.

1. Injection Molding Sequence and Technique

Injection molding machines consist of two basic components: the injection unit and the clamping unit. The material is heated and melted by the injection units, and it is transferred under high pressure into a mold, which is held shut by the clamping unit. A typical operating sequence starts with the feeding of the material by gravity from a hopper, the screw being in the forward position. The rotation of the screw plasticizes the material and conveys it forward in the screw barrel. Accumulation of the molten material at the front of the screw tip forces the screw backwards. After the injection unit is activated, the screw moves forward, pushing the materials from the injection unit, through a nozzle, into the mold. Usually, the injection molding cycle may be considered to consist of three stages: filling, packing, and cooling. During the filling stage, the polymer melt is filled within a short injection time into an empty cavity of the mold, which is fixed by the clamping unit. After filling is completed, more molten polymer is then packed into the cavity at high pressure to compensate for shrinkage as the material cools. Cooling continues after packing is completed. During the cooling stage, the material solidifies and the pressure decreases. The cycle continues until enough rigidity is achieved to permit ejection of the part without damage. The cycle is repeated to produce the next part.

Normally, commercial injection molding involves mold cavities of complex geometry, sometimes with inserts, and a variety of obstructions to direct the flow. The gates in the mold may be located at various positions, and can have a multitude of shapes and dimensions. In some cases, multiple gates are employed to fill a single mold cavity. All these arrangements may have a profound influence on the dynamics of the molding process and on the ultimate characteristics of the molded articles. In particular, some molding arrangements lead to the formation of weld lines in regions where melt fronts intersect inside the cavity. Unless carefully manipulated and controlled, weld lines can represent very weak areas in the molding, leading to premature failure.

2. Structure Development

While the polymer melt passes through very narrow passages of the injection mold such as sprues, runners, gates, and cavities, polymer chains undergo deformation under the high shear stress and rapid cooling near the mold surface. Thus, injection molded parts produced from semicrystalline polymer melts exhibit a complex distribution of crystalline morphology. To clarify the mechanism of formation of these complex structures in injection molded products, we need to confirm the basic characteristics of polymer melts under shear and elongation, e.g., planar, uniaxial, and biaxial, and to understand the crystallization mechanism. According to the characteristics of the crystallization kinetics of semicrystalline polymers, these polymers can be classified into two categories: fast crystallizing polymers, e.g., polyethylene, polypropylene (PP), polyamide, polyoxymethylene, etc.; and slowly crystallizing polymers, e.g., poly(ethylene terephthalate) (PET), poly(phenylene sulfide), poly(ethylene 2,6-naphthalene dicarboxylate), etc., A typical cross-sectional morphology by optical microscopy of an injection-molded PP part is given in Fig. 1(a). The molded PP parts possess an oriented skin layer and a spherulitic core. Between these two regions, an intermediate morphology of ellipsoidal spherulites is observed.

The general observations described above were confirmed by the studies of Kantz *et al.* (1972) and Katti and Schultz (1982) on PP molded bars, using optical microscopy and wide-angle x-ray scattering (WAXS). In particular, the *c*-axes and *a**-axes were found to be bimodally oriented to the mechanical direction (MD), and the *b*-axes rotated around the MD, in the skin layer; and the *c*-axis and *a**-axis orientation to the MD was weak, and the *b*-axes rotated around the MD, in the core layer. Fujiyama (1995) also reported the crystalline structure of an injection molded polypropylene with filler and some nucleating agents dealing with the effect of molding conditions. The molecular orientation as a function of product depth was analyzed, using WAXS and infra-red spectroscopy. It was found that the sub-skin layer, which means the intermediate layer between skin and core layer, had maximum orientation. This molecular orientation within the sub-skin layer was strongly influenced by molecular weight, and the sub-skin layer indicated low molecular orientation in the case of low molecular weight (Trotignon and Verdu 1987).

Polyoxymethylene molded products have also been investigated, and the observations were similar to those of PP. Decreasing mold wall temperature increased the extent of the oriented regions and decreased the amount of spherulites. Transmission electron microscopy (TEM) and small-angle x-ray scattering (SAXS) examinations of these injection molded parts confirmed the oriented character of the outer layers. The results have been used to develop a morphological model of the molded parts in terms of folded chain lamellae, randomly oriented in the

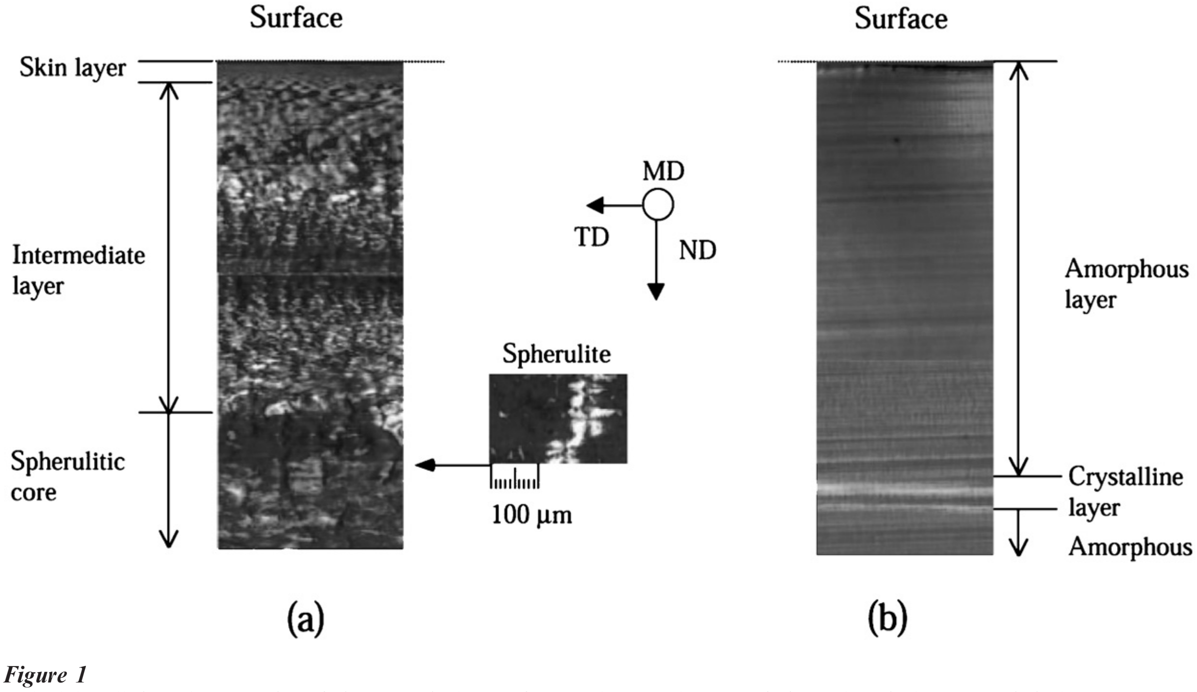


Figure 1
Polarizing light micrographs of the morphology of (a) polypropylene, and (b) poly(ethylene terephthalate) parts: cross-section of rectangular molded parts at 75 mm from the gate. Both samples molded at injection speed of $2.5\,\mathrm{cm^3 s^{-1}}$, and mold temperature in the case of polypropylene and poly(ethylene terephthalate) is 40 and 64 °C, respectively.

center of the pot, and oriented with chains in the flow direction in the outer layers.

The morphology of slowly crystallizing polymers such as poly(phenylene sulfide), poly(ethylene 2,6-naphthalene dicarboxylate), and syndiotactic polystyrene (Ulcer *et al.* 1996) were investigated. These polymers have high glass transition temperature above room temperature, and high flow activation energy that arises from a primary molecular structure containing aromatic groups. A typical cross-sectional morphology of an injection molded PET part is given in Fig. 1(b). Amorphous–crystalline–amorphous multi-layer structures in PET were observed in the gap-wise direction in the case of low injection speed and comparatively low mold temperature. These structural features were found to be very sensitive to cooling rate and shear stress effects.

3. *Modeling*

3.1 General Equation in the Injection Molding Process

Ideally, engineering analysis, mathematical modeling, or computer simulation of the injection molding process should permit the treatment of flow and heat transfer phenomena in complex injection molding cavities. Modern finite element numerical techniques make this possible. Even with the finite element method, the polymers, molds and phenomena involved are so complex that simplification is essential.

One common approach involves the mold network method. According to this method, the complex mold cavity is divided into a number of units of reasonably simple geometry (e.g., rectangular, circular, or cylindrical segments). The solutions for the melt behavior in each of the simpler units are treated separately and are sometimes readily available in graphical, numerical, or analytical forms. To obtain the overall solution for the complete cavity, the solutions for the individual components of the network are coupled and dovetailed to ensure the systematic integration of the results from each segment. Recently, commercial software for the analysis of injection molding process has been developed with the finite element method or the finite difference method. It is useful for optimizing the molding conditions and for aiding in the design of the molds.

The filling stage is concerned with the unsteady-state flow of a hot, compressible, viscoelastic melt into an empty cavity, which is held at a temperature below the solidification temperature of the melt. In this sense, the process can be fully described in terms of the equations of conservation of mass, momentum and energy, coupled with appropriate thermodynamic relationships and a set of constitutive relations which describe the behavior of the material under the influence of stress and thermal fields. In general tensorial formulation, the conservation equations are as follows.

Continuity:

$$\frac{\mathrm{D}\rho}{\mathrm{D}t} = -\rho(\nabla \cdot v) \tag{1}$$

Momentum:

$$\rho\,\frac{\mathrm{D}v}{\mathrm{D}t} = -\nabla P - (\nabla \cdot \tau) + \rho g \tag{2}$$

Energy:

$$\rho C_v\,\frac{\mathrm{D}T}{\mathrm{D}t} = -(\nabla \cdot q) - T\left(\frac{\partial p}{\partial T}\right)\rho(\nabla \cdot v) - (\tau : \nabla v) \tag{3}$$

where ρ is the density, v the velocity, p the pressure, $\boldsymbol{\tau}$ the stress tensor, q the heat flux, T the temperature, and t the time. To obtain a solution of the above equations, each stage of the process is associated with a set of initial and boundary conditions coupled with simplifying assumptions.

In the filling stage, which starts with an empty cavity, a free surface exists at the advancing melt front. In a rectangular cavity, the part filled with the melt is shown in Fig. 2. Near the gate, the flow is radial until the flow front reaches the wall. Away from the gate, the flow is characterized by a dominant axial velocity component. In the front region, the melt, channeling through the center, tends to move towards the cavity wall in what is referred to as the "fountain flow." Although it is possible to analyze

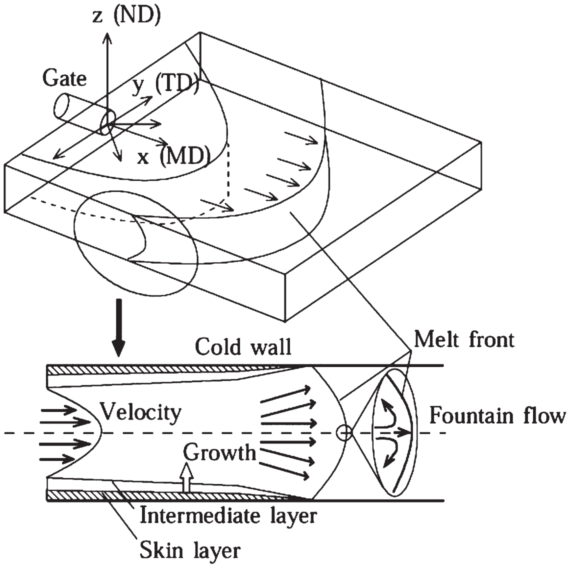


Figure 2
Filling pattern in the melt front of a rectangular injection molding cavity.

the details of the flow structure in the cavity by dealing with the appropriate three-dimensional (3-D) flow equations, it is much simpler to treat the flow regions and to solve the corresponding two-dimensional (2-D) flow equations. In most cases, the flow simulation has employed 2-D solutions by using the Hele–Shaw flow approximation. This approximation may apply to thin mold cavities with large aspect ratios, and in which out-of-plane flows are neglected.

Both finite element and finite difference numerical techniques may be employed to obtain computer solutions of the relevant equations. The former methods are more complex, but they are well suited for dealing with complex mold cavities and the free surface flows at the melt front (Ito *et al.* 1996). An interesting approach for dealing with the fountain flow near the surface is based on the finite difference method in conjunction with the marker-and-cell. Recently, instead of carrying out the full 3-D analysis, a hybrid 2-D/3-D technique was proposed, which solves the 3-D fluid flow equations only in the fountain flow region, while the 2-D Hele–Shaw formulation is employed behind the flow front region. Another approximation is based on the so-called 2.5-D analysis, employing a 2-D flow analysis and a solution to the 3-D energy equation regarding temperature distribution and crystallization effects using Boundary Fitted Curvilinear Coordinate transformation. Hsiung *et al.* (1996) developed the model using a Lagrangian approach to simulate the stress distribution along the gap-wise direction inside the injection molding.

There are a variety of constitutive rheological equations that may be employed to describe the viscous or viscoelastic behavior of the polymer melt during the process. The most commonly employed rheological models in the injection molding process include the following: Newtonian liquid, power law model, second-order fluid, Leonov equation, and integral constitutive K-BKZ equation (Chang and Chiou 1994). Whereas some of the models describe only the viscous behavior of the fluid, other models reflect some of the viscoelastic characteristics of the material.

The approach for the simulation of the packing stage is similar to that for the filling stage. For the packing stage, the deceleration of the flow reinforces the importance of the unsteady-state terms, and should not be neglected. During packing, the contribution of the unsteady-state terms will tend to be larger than that of the convective terms in the equation.

The dominant feature in the packing and cooling stage is the compressibility of the polymer melt. This emphasizes the need for selecting an appropriate equation of state to describe the pressure–volume–temperature (PVT) relationships for the melt. A number of equations of state have been employed in conjunction with the processing of polymer melts, the majority of which are based on the van der Waals equation of state.

After the filling and packing of the cavity, the gate freezes and no additional material enters the cavity. Since fluid motion is negligible or absent during this stage, the problem is reduced to the transient heat transfer by conduction from the polymer melt to the mold walls. The convection in the 2-D stream-wise direction and thermal conduction in the gap-wise direction are considered in the energy equation. Also the generation of heat due to shear flow is taken into account.

$$\rho C_p\left(\frac{\partial T}{\partial t}+u\frac{\partial T}{\partial x}+v\frac{\partial T}{\partial y}\right) = \frac{\partial}{\partial z}\left(\lambda\frac{\partial T}{\partial z}\right)+\tau_{zx}\frac{\partial u}{\partial z}+\tau_{zy}\frac{\partial v}{\partial z} \tag{4}$$

where x and y are the stream-wise direction, z the thickness direction, and t the time. u is the velocity in the x-direction and v the velocity in the y-direction, respectively. τ is the shear stress, T the temperature, with ρ, C_p, and λ representing the density, specific heat, and thermal conductivity.

In order to allow for the effect of latent heat of crystallization, the latent heat is included in the specific heat as follows:

$$C_p = C_p' + \Delta H\,\frac{\mathrm{d}x}{\mathrm{d}t} \tag{5}$$

where C_p is the effective specific heat capacity of the material, C_p' is the specific heat of the material at the same temperature in the absence of crystallization, ΔH is the latent heat of crystallization, and x is the weight fraction crystallinity in the system.

3.2 Crystallization Kinetics using Crystallization Rate Constant

Simulation of crystallization during injection molding has been performed (Kamal and Lafleur 1986, Hsiung *et al.* 1996, Titomanlio *et al.* 1995, Isayev *et al.* 1995), where the form of the crystallization kinetics is the generalized Avrami equation by Nakamura *et al.* (1972). This model is based on the assumption of isokinetic conditions, and it describes the crystallization process by a single parameter, i.e., crystallization rate constant.

$$\frac{x(t)}{x_\infty} = 1-\exp\left[-\left(\int_0^t K(T(t)\,\mathrm{d}t\right)^n\right] \tag{6}$$

where $x(t)$ is crystallinity, x_∞ is the final crystallinity, n is the Avrami index determined from isothermal experiments, and $K(T)$ $(=[k(T)]^{1/n})$ is connected with the crystallization rate constant of isothermal crystallization $k(T)$. A crystallization kinetics model that took into account the effect of shear stress on crystallization rate constant and induction time has been proposed, and the distributions of crystallinity along the stream-wise and gap-wise predicted; good agreement with experimental values was observed (Hsiung *et al.* 1996). Another model of shear-induced crystallization kinetics in the injection molding process was proposed by Isayev *et al.* (1995), which took into account the shear effect on the rheological parameters, i.e., shear rate and relaxation time. This simulation model is based on the theory of Eder and Janeschitz-Kriegl *et al.* (1990), which assumes that the locations of nucleation are created by the flow. Friedl and McCaffrey (1991) simulated the crystallinity distribution of polyoxymethylene across a rectangular part, and crystallinity profile through a section, where the macrokinetics approach of Malkin *et al.* (1984) was used to model the crystallization behavior during the injection molding process. In addition, Titomanlio *et al.* (1995) proposed a model of crystallization kinetics which took into account the effect of pressure on the equilibrium crystallinity, and showed that they were able to increase the accuracy of the prediction of the crystallinity distribution by considering the pressure effect.

3.3 Crystallization Kinetics Considering Nucleation and Crystal Growth Process

The actual crystallization process, however, consists of two elementary processes: nucleation of crystals and their growth. These two processes are independently influenced by temperature, pressure and shear stress. In order to simulate the crystal morphology and more realistic crystallization behavior during the molding process, crystallization kinetics should be described using separate terms for the nucleation rate and the crystal growth rate.

Ishizuka and Koyama (1977) have shown that the crystallization behavior of a running filament is well reproduced by using generalized Avrami kinetics, that take into account the temperature dependence of the nucleation rate and the crystal growth rate. In the case of a heterogeneous nucleation process, the crystallinity $x(t)$ is given by:

$$-\ln\left(1-\frac{x(t)}{x_\infty}\right) = \frac{1}{x_\infty}\frac{\rho_c}{\rho_l}\left[\int_0^t \frac{d\bar{N}}{d\tau}\, v(t,\ \tau)\, d\tau + \bar{N}(0)v(t,\ 0)\right] \quad (7)$$

where

$$v(t,\ \tau) = k_f\left[\int_\tau^t G(u)\, du\right]^m \quad (8)$$

Here x_∞ is the final crystallinity, $G(u)$ the crystal growth rate at time u, and $N(t)$ the number of nuclei at time t. m is the dimension of the growth of the crystal, and k_f is the corresponding shape factor. In the case of three-dimensional crystal growth, $m = 3$ and $k_f = 4\pi/3$ are adapted. ρ_c and ρ_l are the density of the crystalline and liquid phases, respectively. The effects of orientation were also considered through the effect of shear stress on the melting temperature in the nucleation rate and crystal growth rate, which was based on work by Kobayashi and Nagasawa (1970).

Crystallization behavior in the injection molding process has been estimated using crystallization kinetics with the density of nuclei and crystal growth rate by Okamoto *et al.* (1995) and Ito *et al.* (1996). They considered the effect of shear stress on the crystallization kinetics, experimentally or theoretically. It is generally considered that the crystallization in large products starts at the filling stage, where the polymer is subjected to rapid cooling, high pressure, and high shear stress. Therefore it is necessary to consider the influence of pressure on the crystallization behavior, because the pressure in the injection molding process of large products often reaches more than 150 MPa during the filling stage, and the high pressure accelerates the crystallization. Simulation of crystallization at high pressure under non-isothermal conditions was carried out, and these simulation values agreed well with experimental values, including the effect of pressure on the melting temperature and glass transition temperature on nucleation and crystal growth rates. Later, Ito *et al.* (1996) described in more detail the effects of pressure and shear stress on the crystallization behaviors during injection molding. They demonstrated that the model developed is a quantitative

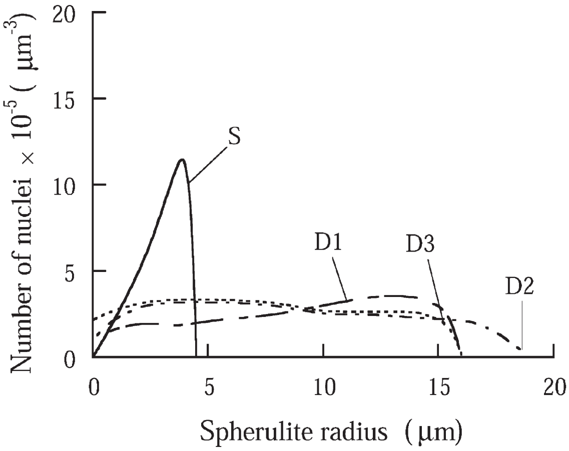


Figure 3
Gap-wise distribution of predicted spherulite size of polypropylene part at 0.5 mm from the gate. Sample molded at injection speed of 22.2 cm^{3}s^{-1} and mold temperature of 0 °C. S: surface layer; D1: 0.2 mm from the surface; D2: 0.4 mm from the surface; D3: center layer.

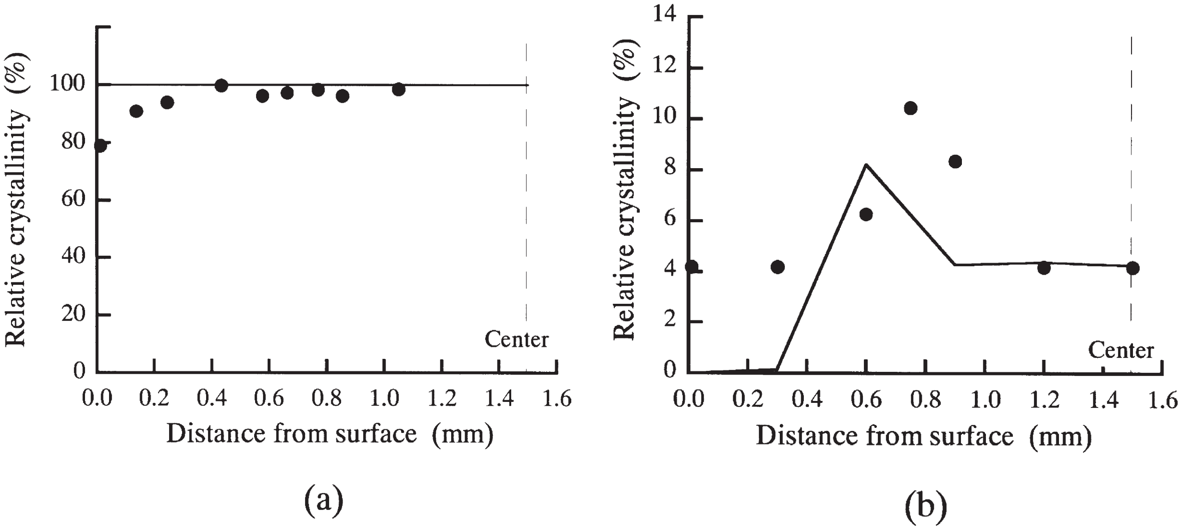


Figure 4
Gap-wise distribution of predicted (line) and measured (symbol) relative crystallinity of (a) polypropylene, and (b) poly(ethylene terephthalate) parts at 50 mm from the gate. Sample molding conditions are indicated in caption for Fig. 1.

phenomenological model that can predict not only the crystallinity distribution, but also the spherulite size distribution in stream-wise and gap-wise directions. The predicted spherulite distribution in the gap-wise direction of injection molded PP part is shown in Fig. 3. Moreover, in order to accurately predict the flow and crystallization phenomena in the injection molding process, material physical properties—such as density, thermal conductivity, and specific heat—should be taken into account, as well as the crystallinity dependence of these material properties. They also linked the simulation model between crystallization and the flow with the effect of crystallization on material physical properties. Figure 4 shows the distribution of relative crystallinity in the gap-wise direction of injection molded PP and PET parts. The gap-wise distribution of relative crystallinity was in good agreement with experimental values.

Bibliography

Chang R -Y, Chiou S -Y 1994 Integral constitutive model (K-BKZ) to describe viscoelastic flow in injection molding. *Int. Polym. Proc.* **IX**, 365–72

Eder G, Janeschitz-Kriegl H, Liedauer S 1990 Crystallization processes in quiescent and moving polymer melts under heat transfer conditions. *Prog. Polym. Sci.* **15**, 629–714

Friedl C F, McCaffrey N J 1991 Crystallisation prediction in injection molding. *Soc. Plast. Eng. ANTEC'91 Tech. Pap.*, pp. 330–2

Fujiyama M 1995 Structures and properties of injection moldings of β-crystal nucleator-added polypropylenes. *Int. Polym. Proc.* **X**, 172–8

Hsiung C M, Cakmak M, Ulcer Y 1996 A structure-oriented model to simulate the shear-induced crystallization in injection molded polymers: a Lagrangian approach. *Polymer* **37**, 4555–71

Isayev A I, Chan T W, Gmerek M, Shimojo K 1995 Injection molding of semicrystalline polymers—II: Modeling and experiments. *J. Appl. Polym. Sci.* **55**, 821–38

Ishizuka O, Koyama K 1977 Crystallization of running filament in melt spinning of polypropylene. *Polymer* **18**, 913–8

Ito H, Minagawa K, Takimoto J, Tada K, Koyama K 1996 Effect of pressure and shear stress on crystallization behaviors in injection molding. *Int. Polym. Proc.* **XI**, 363–8

Kamal M R, Lafleur P G 1986 Computer simulation of injection mold filling for viscoelastic melts with fountain flow. *Polym. Eng. Sci.* **26**, 190–6

Katti S S, Schultz J M 1982 The microstructure of injection-molded semicrystalline polymers: a review. *Polym. Eng. Sci.* **22**, 1001–17

Kantz M R, Newman H D Jr., Stigale F H 1972 The skin-core morphology and structure-property relationships in injection-molded polypropylene. *J. Appl. Polym. Sci.* **16**, 1249–60

Kobayashi K, Nagasawa T 1970 Crystallization of sheared polymer melts. *J. Macromol. Sci.—Phys.* **B4**, 331–45

Malkin A Y, Beghishev V P, Keapin I A, Bolgov S A 1984 General treatment of polymer crystallization kinetics—I: A new macrokinetic equation and its experimental verification. *Polym. Eng. Sci.* **24**, 1396–401

Nakamura K, Watanabe T, Katayama K, Amano T 1972 Nonisothermal crystallization of polymers—I: Relationship between crystallization temperature, crystallinity, and cooling conditions. *J. Appl. Polym. Sci.* **16**, 1077–91

Okamoto M, Shinoda Y, Kinami N, Okuyama T 1995 Non-isothermal crystallization of poly(ethylene terephthalate) and its blends in the injection-molding process. *J. Appl. Polym. Sci.* **57**, 1055–61

Spencer R S, Gilmore G D 1951 Some flow phenomena in the injection molding of polystyrene. *J. Coll. Sci.* **6**, 118–32
Titomanlio G, Speranza V, Brucato V 1995 On the simulation of the thermoplastic injection molding process. *Int. Polym. Proc.* **X**, 55–61
Trotignon J -P, Verdu J 1987 Skin-core structure: fatigue behavior relationships for injection-molded parts of polypropylene—I: Influence of molecular weight and injection conditions on the morphology. *J. Appl. Polym. Sci.* **34**, 1–18
Ulcer Y, Cakmak M, Miao J, Hsiung C M 1996 Structural gradients developed in injection-molded syndiotactic polystyrene. *J. Appl. Polym. Sci.* **60**, 669–91

H. Ito
Tokyo Institute of Technology, Japan

Intermetallics: Crystal Structures

The search for new high-temperature structural materials has stimulated much interest in ordered intermetallic compounds (Westbrook and Fleischer 1995, Williams 1997). Alloys based on the ordered intermetallic compounds constitute a unique class of metallic material that form a long-range ordered crystal structure (as opposed to a random solid solution) below a certain critical temperature, generally referred to as the critical ordering temperature (T_c). For most of the aluminides of nickel, iron, and titanium this T_c temperature is close to (usually above) the melting temperature of an intermetallic. A notable exception is Ti_3Al, which disorders at 1100 °C but has a melting point around 1600°C. Iron aluminide of the Fe_3Al type passes through two ordered structures ($D0_3$ and B2) before becoming disordered (Vedula 1995). It must be pointed out that a perfect ordered crystallographic structure of an intermetallic compound exists only at the exact stoichiometry corresponding to its stoichiometric formula. However, quite a number of intermetallics, including the aluminides of nickel, iron, and titanium, exist over a range of compositions, but the degree of order decreases as the deviation from stoichiometry increases. Additional alloying elements can be incorporated without losing the ordered crystallographic structures. The possibility of alloying the ordered structures of intermetallic compounds opens up new avenues for development of novel high-temperature alloys to improve or optimize properties for specific applications. Therefore, a thorough understanding of how atoms are arranged in an ordered fashion in the crystallographic lattices of intermetallic compounds is a major prerequisite for further development of novel intermetallic-based high-temperature alloys and understanding their mechanical, physical, and chemical properties. This article presents a comprehensive view of the crystal structures of intermetallic compounds, which have some potential as structural materials for commercial applications, mainly at elevated temperatures and hostile environments.

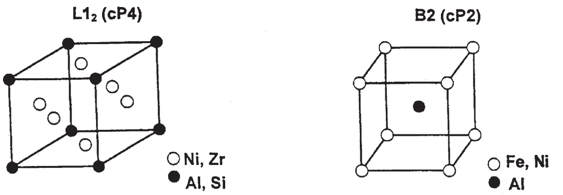
Ordered intermetallics based on aluminides and silicides constitute a major group of intermetallics, which has been extensively researched since the early 1990s. Another important group of intermetallics are the Laves phases. They represent a huge group of binary intermetallic compounds, many of which have good combinations of high melting point, low density, and good oxidation resistance. Many ternary elements and, in particular, refractory elements, have large solid solubility in Laves phases for a potential improvement of mechanical properties. Last but not least, the crystal structures of some non-commercial intermetallic compounds (important from the standpoint of fundamental studies) will also be discussed. Traditionally, the crystallographic lattices of intermetallic compounds have been quoted according to the *Strukturbericht* designation. However, because this system cannot be conveniently and systematically expanded to cover the large variety of crystal structures currently being encountered, the system has gradually fallen into disuse. Nowadays, the system of *Pearson symbols* (Villars and Calvert 1991) has become more widely used. However, because the *Strukturbericht* designations are still quite popular, in the present article each crystallographic lattice of an intermetallic compound will be distinguished by the *Strukturbericht* designation and accompanied by the corresponding Pearson symbol in parentheses.

1. *Aluminides and Silicides*

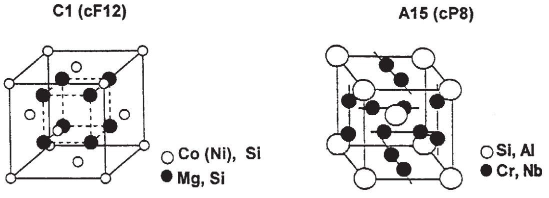
1.1 Generic Cubic Crystal Structures

Figure 1 shows hard-sphere models of the generic crystallographic unit cells of cubic aluminides and silicides. The term "generic" means that these crystal structures occur in aluminides and silicides naturally, for example upon congruent solidification from the melt without any change in composition.

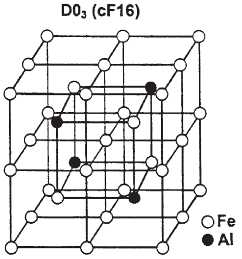
Intermetallic compounds such as Ni_3Al, Zr_3Al, and Ni_3Si have an $L1_2$ (cP4) crystal structure (Fig. 1(a)), a derivative of the face-centered cubic (fcc) crystal structure. This unit cell contains four atoms, i.e., the same number as fcc. The nickel (Ni_3Al, Ni_3Si) or zirconium (Zr_3Al) atoms occupy face-centered positions and the aluminum or silicon atoms occupy the corners of the unit cell. Both NiAl and FeAl possess a B2 (cP2) crystal structure a derivative of the body-centered cubic (bcc) crystal structure (Fig. 1(b)). The unit cell contains two atoms, i.e., the same number as bcc. In reality, it consists of two interpenetrating primitive cubic cells and is sometimes referred to as the *CsCl* type. The aluminum atoms occupy the cube corners of one sublattice and the nickel atoms occupy the cube corners of the second sublattice.



(a) Ni_3Al, Ni_3Si, Zr_3Al (b) NiAl, FeAl



(c) Mg_2Si, $CoSi_2$, $NiSi_2$ (d) Cr_3Si, Nb_3Al



(e) Fe_3Al

Figure 1
Generic cubic lattices of aluminides and silicides. (a) $L1_2$ (cP4), (b) B2 (cP2), (c) C1 (cF12), (d) A15 (cP8), and (e) $D0_3$ (cF16).

Silicides such as Mg_2Si, Co_2Si, and Ni_2Si possess a more complex C1 (cF12) crystal structure with 12 atoms per unit cell (Fig. 1(c)). The archetype of this structure is the ionic ceramic compound CaF_2 (calcium fluorite). This ionic structure can be viewed as built on an fcc lattice of small cations (Ca^{2+}) with eight bigger anions (F^-) forming a simple cubic sublattice (at $\frac{1}{4}\ \frac{1}{4}\ \frac{1}{4}$ etc. positions) inside the fcc unit cell. There are ($4Ca^{2+}+8F^-$) ions per unit cell. Such a crystal structure is termed a *fluorite structure*. If in some ionic compounds the fcc lattice is built on large anions, and smaller cations are located inside the fcc cell, then the C1 crystal structure is termed an *anti-fluorite structure*. From the point of view of atomic size, Mg_2Si (and its companion intermetallics Mg_2Ge, Mg_2Sn, and Mg_2Pb) is regarded as fluorite type whereas $CoSi_2$ and $NiSi_2$ are regarded as anti-fluorite type.

The A15 (cP8) cubic structure with A_3B stoichiometry (Cr_3Si and Nb_3Al) and eight atoms per unit cell consists of orthogonal chains of "A" (Cr or Nb) atoms along $\langle 100 \rangle$ directions criss-crossing the bcc arrangement of "B" (Si or Al) atoms (Fig. 1(d)). The A15 is considered the simplest of the so-called *topologically close-packed* (TCP) structures (Khantha *et al.* 1989).

Finally, Fe_3Al has the ordered cubic $D0_3$ (cF16) crystal structure (Fig. 1(e)). This unit cell contains eight bcc-type subcells and it may be thought of as being composed of four interpenetrating fcc lattices (notice that according to the system of Pearson symbols the $D0_3$ lattice is classified as cubic face-centered—cF). In each subcell the iron atoms occupy corners and as such each of them is shared with eight neighboring subcells, i.e., one eighth of the iron atom per subcell. Since there are eight subcells, the corner atoms provide eight iron atoms per unit cell. In addition, four iron atoms occupy the centers of four subcells, so altogether there are twelve iron atoms in the $D0_3$ unit cell. Including the four aluminum atoms, which also occupy the centers of four subcells, the total number of atoms per unit cell comes to sixteen. Fe_3Al transforms from the $D0_3$ structure to a defective ordered cubic B2 (cP2) crystal structure at temperatures above the critical ordering temperature T_c of ~814K (Vedula 1995).

1.2 Cubic Aluminides Stabilized by Alloying

Low-density and highly oxidation-resistant titanium and zirconium trialuminide intermetallics, with respective stoichiometric formulas Al_3Ti and Al_3Zr (trialuminides), crystallize in tetragonal $D0_{22}$ (tI8) and $D0_{23}$ (tI16) crystallographic lattices, respectively (Fig. 2(b),(c)). Their crystallographic structures are closely related to the $L1_2$ structure (Fig. 2(a)). They can be essentially derived from the former by introducing an antiphase boundary (APB) with a displacement vector of $a/2\ \langle 100 \rangle$ on every (001) plane in $D0_{22}$ and every second (001) in the $D0_{23}$ lattice (Fig. 2(a),(b),(c)). It must, however, be pointed out that a close examination of the close-packed planes and their respective order in the $L1_2$, $D0_{22}$, and $D0_{23}$ structures reveals that the order in these planes is distinctly different, being triangular (T) (Fig. 2(a)), rectangular (R) (Fig. 2(b)), and mixed (T-R) (Fig. 2(c)) in the $L1_2$, $D0_{22}$, and $D0_{23}$ structures, respectively. A close relationship between tetragonal and cubic structures in Fig. 2 allows transformation of tetragonal $D0_{22}$ Al_3Ti into the $L1_2$ structure by alloying with approximately 8–10 at% copper, nickel, and zinc (Schubert *et al.* 1964a, Raman and Schubert 1965), chromium and manganese (Zhang *et al.* 1990), iron (Kumar and Pickens 1988), and palladium (Powers and Wert 1990). Similar transformation of the low-symmetry $D0_{23}$ Al_3Zr into the $L1_2$ structure can be achieved by alloying with copper, nickel, and zinc (Schubert *et al.* 1964a, b), chromium and iron (Schneibel and Porter 1989), and manganese (Virk and Varin 1991, 1992). There is a general consensus

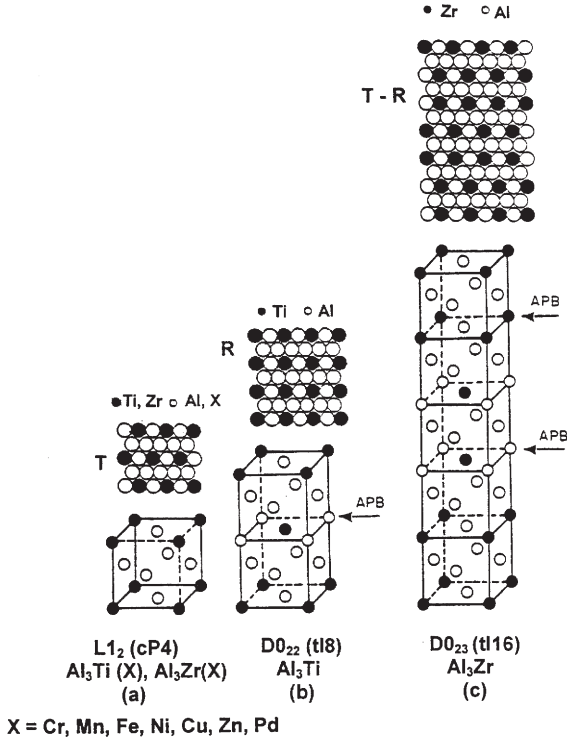


Figure 2
The $L1_2$ structure in titanium and zirconium trialuminides, which can be stabilized by alloying of tetragonal $D0_{22}$ Al_3Ti and $D0_{23}$ Al_3Zr with Cr, Mn, Fe, Ni, Cu, Zn, and Pd. A hard-sphere model of the close-packed planes in each crystal structure is also shown. (a) Stabilized $L1_2$ structure of Al_3Ti (X) where X is the appropriate alloying element; triangular (T) ordering, (b) $D0_{22}$ Al_3Ti; rectangular (R) ordering, and (c) $D0_{23}$ Al_3Zr; mixed triangular–rectangular (T-R) ordering.

that the alloying element atoms substitute mostly for the aluminum atoms. It is to be pointed out that the first case of systematic alloying-induced transformation to the $L1_2$ structure was reported for $D0_{22}$-structured Ni_3V and hexagonal $D0_{19}$ (hP24)-structured Co_3V by Liu (1984) (so-called *long range-ordered* (LRO) alloys). Both low-symmetry intermetallics are transformed to the $L1_2$ structure by partially substituting iron for cobalt and nickel, respectively. The crystallographic transformation to the $L1_2$ structure in the LRO alloys results in a dramatic increase in tensile ductility at room temperature as opposed to the same transformation in Al_3Ti, which does not lead to any improvement in tensile ductility.

1.3 Noncubic Aluminides and Silicides

Commercially important aluminides and silicides such as $D0_{19}$ (hP8) Ti_3Al, $L1_0$ (tP4) TiAl, $C11_b$ (tI6) $MoSi_2$, tetragonal tP32 (no *Strukturbericht* designation; PTi_3 prototype) Nb_3Si, and its counterpart the tetragonal $D8_l$(tI32) α-Nb_5Si_3 (or $D8_m$ (tI32) β-Nb_5Si_3), do not possess cubic structures. Nb_5Si_3 belongs to a wider group of intermetallic compounds, so-called *5:3 silicides*, having a general stoichiometric formula M_5Si_3 ($M = Nb$, Ta, Mo, V, Ti, Zr, W). These compounds typically crystallize in either a tetragonal $D8_l$ (tI32) (e.g., Nb_5Si_3, Ta_5Si_3, Mo_5Si_3, V_5Si_3, W_5Si_3) or a hexagonal $D8_8$ (hP16) structure (e.g., Ti_5Si_3). However, the tetragonal M_5Si_3 can also crystallize with the $D8_8$ structure if they contain small amounts of interstitial impurities, in particular carbon (Sauthoff 1996). Figure 3 shows the hard-sphere models of selected noncubic crystallographic structures.

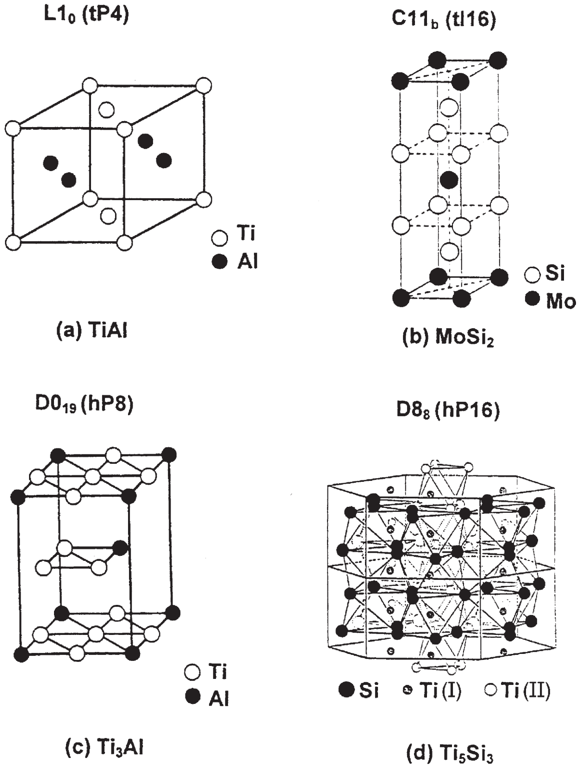


Figure 3
Hard-sphere models of crystalline unit cells of non-cubic aluminides and silicides. (a) $L1_0$ (tP4), (b) $C11_b$ (tI16), (c) $D0_{19}$ (hP8), and (d) $D8_8$ (hP16).

The crystallographic unit cell of TiAl (Fig. 3(a)) is commonly referred to as a face-centered tetragonal (fct). The tetragonality is very small ($c/a = 1 \cdot 02$). The structure of $MoSi_2$-type phases (Fig. 3(b)) has been described as a superstructure of the bcc or CsCl structure with three subcells stacked along the [001] direction. The c/a ratio for $MoSi_2$ is 2.45. Ti_3Al has a hexagonal structure (Fig. 3(c)) with eight atoms per unit cell. Finally, the hexagonal structure of Ti_5Si_3 is

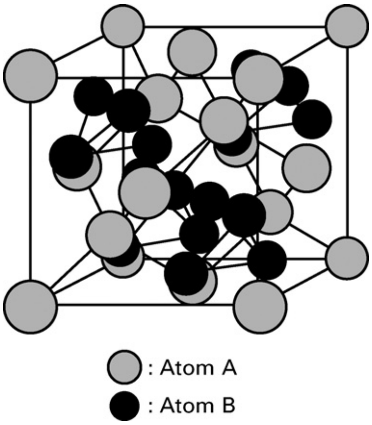


Figure 4
The C15 unit cell characteristic of Laves phases.

built up by two atomic species which occupy three different crystallographic positions (Fig. 3(d)). The titanium(I) atoms are arranged in chains parallel to each other octahedrally surrounded by silicon atoms. In contrast, the titanium(II) atoms form columns of octahedrons connected along the *c*-axis. The octahedral interstices can be partially or completely filled by interstitials.

2. *Laves Phases*

The binary Laves phases with an AB_2 composition belong to the above-mentioned *topologically close-packed* (TCP) structures. In the group of TCP intermetallics, the Laves phases constitute the single largest group. They crystallize in the hexagonal C14 (hP12; $MgZn_2$ prototype), the cubic C15 (cF24; Cu_2Mg prototype) (Fig. 4), or the dihexagonal C36 (hP24; $MgNi_2$ prototype) structures. The C14, C15, and C36 crystal structures differ only by the particular stacking of the same two-layered structural units, which allows structure transformations between these structures and twinning by "synchroshear" (Hazzledine *et al.* 1993). The stability of the three crystal structures is controlled by both the atomic size ratio of the A atoms and B atoms and by the valence electron concentration of the Laves phase (Sauthoff 1996). Some Laves phases have been regarded as promising for both functional and structural applications, such as superconducting C15 (Hf, Zr)V_2, magnetic C15 TFe_2 (where $T = Ti$, Zr, Hf, Nb, and Mo), and materials for hydrogen storage C15 ZrV_2, $ZrCr_2$, and C14 $ZrMn_2$ materials (Sauthoff 1996).

Bibliography

Hazzledine P M, Kumar K S, Miracle D B, Jackson A G 1993 Synchroshear of Laves phases. *Mater. Res. Soc. Symp. Proc.* **288**, 591–6

Khantha M, Vitek V, Pope D P 1989 Twinning in A15 compounds. *Mat. Res. Soc. Symp. Proc.* **133**, 99–104

Kumar K S, Pickens J R 1988 Ternary low-density cubic $L1_2$ aluminides. In: Kim Y W, Griffith W M (eds.) *Dispersion Strengthened Aluminum Alloys.* TMS, Warrendale, PA, pp. 763–86

Liu C T 1984 Physical metallurgy and mechanical properties of ductile ordered alloys $(Fe, Co, Ni)_3V$. *Int. Met. Rev.* **29**, 168–94

Powers W O, Wert J A 1990 Microstructure, deformation and fracture characteristics of an $Al_{67}Pd_8Ti_{25}$ intermetallic alloy. *Metall. Trans. A* **21A**, 145–51

Raman A, Schubert K 1965 Über den Aufbau einiger zu $TiAl_3$ verwandter Legierungsreihen III. Untersuchungen in einigen Ti–Ni–Al und Ti–Cu–Al Systemen. *Z. Metallkd.* **56**, 99–101

Sauthoff G 1996 Intermetallics. In: Cahn R W, Haasen P, Kramer E J (eds.) *Materials Science and Technology*. VCH, Weinheim, Germany, Vol. 8, pp. 646–766

Schneibel J H, Porter W D 1989 Microstructure and mechanical properties of $L1_2$-structure alloys based on Al_3Zr. In: Liu C T, Taub A I, Stoloff N S, Koch C C (eds). Mater. Res. Soc. Symp. Proc. 133, 335–40

Schubert K, Meissner H G, Raman A, Rossteutscher W 1964a Einige Strukturdaten metallischer Phasen (9). *Naturwissenschaften* **51**, 287

Schubert K, Raman A, Rossteutscher W 1964b Einige Strukturdaten metallischer Phasen (10). *Naturwissenschaften* **51**, 507

Vedula K 1995 FeAl and Fe_3Al. In: Westbrook J H, Fleischer R L (eds.) *Intermetallic Compounds—Principles and Practice.* Wiley, Chichester, UK, Vol. 2, pp. 199–209

Villars P, Calvert L D 1991 *Pearson's Handbook of Crystallographic Data for Intermetallic Phases.* ASM International, Metals Park, OH

Virk I S, Varin R A 1991 Development of structure and porosity in cast Al_5CuZr_2 and $Al_{66}Mn_9Zr_{25}$ intermetallic compounds. *Metall. Trans. A* **22A**, 2545–52

Virk I S, Varin R A 1992 Mechanical behavior and microcracking of cubic ternary zirconium trialuminides. *Metall. Trans. A* **23A**, 617–25

Westbrook J H, Fleischer R L (eds.) 1995 *Intermetallic Compounds—Principles and Practice.* Wiley, Chichester, UK

Williams J C 1997 Intermetallics for structural applications: potential, reality and the road ahead. In: Nathal M V, Darolia R, Liu C T, Martin P L, Miracle D B, Wagner R, Yamaguchi M (eds.) *Structural Intermetallics 1997.* TMS, Warrendale, PA, pp. 3–8

Zhang S, Nic J P, Mikkola D E 1990 New cubic phases formed by alloying Al_3Ti with Mn and Cr. *Scr. Metall. Mater.* **24**, 57–62

R. A. Varin
University of Waterloo, Ontario, Canada

Intermetallics: Laves Phases

The class of intermetallics referred to as "Laves phases" was first discovered by James B. Friauf (1927). Friauf's x-ray diffraction investigations focused upon the crystallographic forms of $MgCu_2$ and $MgZn_2$. Following that discovery, in the 1930s Fritz

Laves (e.g., Laves and Witte 1936) initiated a comprehensive investigation of the characteristics and features of these materials. He ultimately reviewed this work in articles published in 1954 and 1967. As a consequence of the insight and body of knowledge that Laves supplied, these intermetallic phases commonly bear his name. Subsequently, insightful interpretations of the crystal structure by Komura (1962), as well as the occurrence of stacking variants in the intermetallic phase (Komura 1977), have furthered the understanding of this class of materials. Accordingly, owing to the contributions of these three primary investigators, literature references to the intermetallic phases include FLK (Friauf–Laves–Komura) phases. This article will review the characteristics, properties, and applications of this class of materials.

1. Crystallographic Features

1.1 Crystal Structure

Laves phases are topologically close-packed (t.c.p.), ordered, intermetallic compounds with the approximate formula AB_2. Most of these intermetallics crystallize in one of the three crystal structure polytypes (Strukturbericht designations of C14, C15, and C36). The A atom is the larger of the two elements that comprise the binary structure. Overall, the smaller B atoms are coordinated tetrahedra that are linked corner-to-corner (Fig. 1). The A atoms fit into the interstices enclosed by the B-atom tetrahedra and are coordinated into a wurtzite structure for the C14 structure and into a diamond structure for the C15 structure.

Individually, the C14 structure ($MgZn_2$ prototype) has $P6_3/mmc$ symmetry in Hermann–Maugin symbols. The structure is classified as hP12 in Pearson notation. The C15 structure ($MgCu_2$ prototype) has the structure notation of cF24 (Pearson notation) and the Hermann–Maugin space group is $Fd\bar{3}m$. The C36 structure ($MgNi_2$ prototype) is a double hexagonal representation of the C14 phase, and therefore has the same symmetry ($P6_3/mmc$), but its structure notation is hP24. The C36 phase is structurally an intermediate stacking sequence between the C14 structure and C15 structure.

All Laves phases (including all polytypes, crystal defects, etc.) can be constructed from one of six four-layer fundamental stacking schemes of close-packed planes shown in Fig. 2. The elementary stacking units of the four interpenetrating atomic layers are defined for the C15 phase by …XYZXYZ…, for the C14 phase by …XY′XY′…, and the for C36 phase by …XY′X′ZXY′X′Z…. In addition, a variety of other intermediate stacking sequences (up to 21 layers) have been defined. The three main polytypes have 12, 24, and 24 atoms per unit cell, respectively, and the structures vary only in the stacking sequence of atom planes (similar to disordered f.c.c. and h.c.p. metals). The space filling and coordination are unchanged between the structures. Based on a hard-sphere model, the atomic packing factor is 0.71 when $r_A/r_B = \sqrt{3/2} \approx 1.225$, where r_A and r_B are the radii of the metal atoms in the ordered phase.

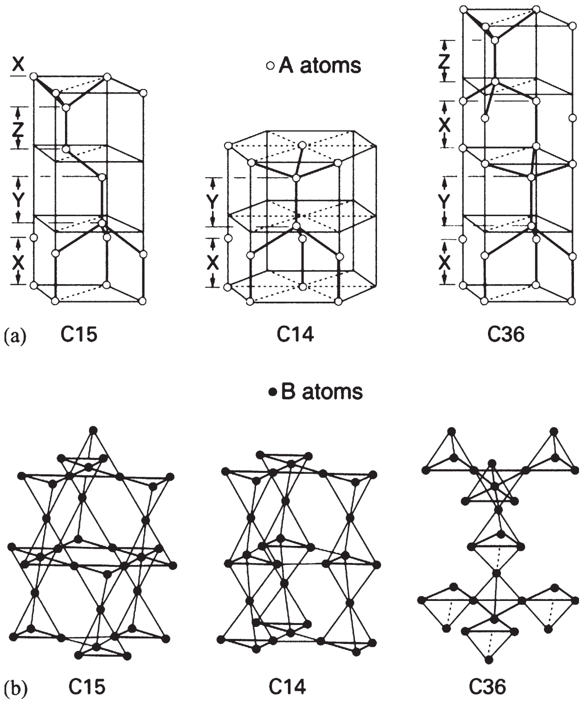


Figure 1
Distribution of (a) the A atoms and stacking of layers and (b) the B atoms and stacking of tetrahedra in Laves phases (after Barrett and Masalski 1980).

1.2 Bond Distances

The basic unit cells for the C14 and C15 structures are shown in Figs. 3 and 4. A cross-section of each unit cell is also illustrated. From these figures, the A atoms have a coordination number of 16 (4A + 12B atoms) and the B atoms have a coordination of 12 (6A + 6B atoms). The geometry of bond distances (assuming hard-sphere models) are demonstrative in understanding the nature of Laves phases. For the C14 bond distances, the respective A–A, B–B, and A–B distances are not all equidistant unless the axial ratio c/a is exactly $\sqrt{(8/3)}$. At $c/a = \sqrt{(8/3)}$, d_A/d_B (where d is the diameter of the atom in the intermetallic unit cell) is equal to $\sqrt{(3/2)}$ (or ≈ 1.225) as discussed above. It is important to note that based upon the cross-sections of the unit cells shown in Figs. 3 and 4, the diameters for the A and B atoms are actually the bond distances. Following Berry and Raynor (1953) for nonideal axial

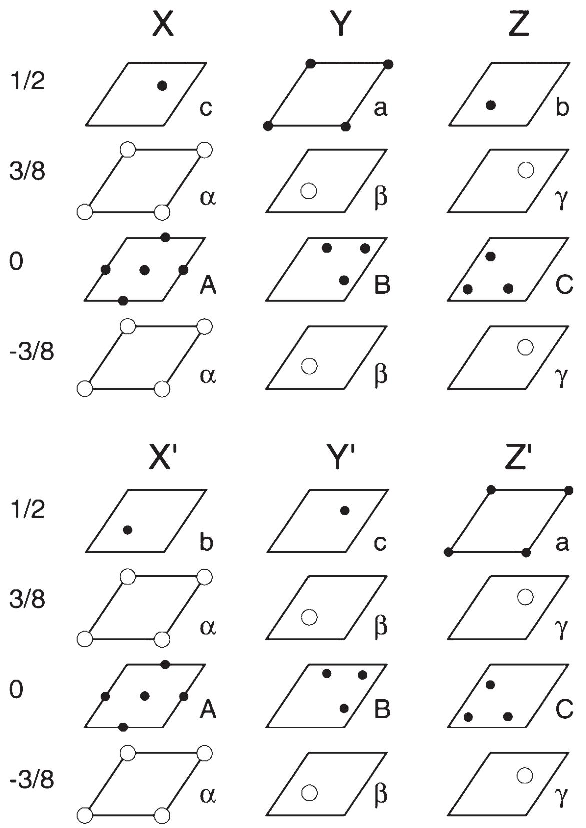


Figure 2
Six fundamental four-layer stacking schemes of the close-packed direction of the Laves phases (after Komora 1962).

ratios, two distances can be defined for the A and B atoms, The A distances are $d_{A1}=c(1/2-2z)$ and $d_{A2}=[a^2/3\ +\ (2cz)^2]^{1/2}$ where z is a parameter that defines the relative position of A atoms above or below the basal plane. The value of z is dependent upon the specific alloy phase but is very near 1/16. The B distances can be defined as follows: $d_{B1}=(1+3x)a$ and $d_{B2}=(3x^2a^2+c^2/16)^{1/2}$, where x corresponds to the B atom position in the lattice with respect to the a lattice parameter and is approximately $-1/6$. In addition to the A and B distances, the A–B distances can be defined as: $d_{AB1}=(a^2/3+c^2z^2)^{1/2}$, $d_{AB2}=[a^2(1/3+x+3x^2)+c^2(1/4-z)^2]^{1/2}$, and $d_{AB3}=[3a^2(1/3+x)^2\ +c^2(z+1/4)^2]^{1/2}$. The bond distances for the C15 structure type are defined as follows: $d_A=(a\sqrt{3})/4$, $d_B=(a\sqrt{2})/4$, and $d_{AB}=(a\sqrt{11})/8$.

2. *Abundance of Laves Phases*

Laves phases form the most abundant class of binary intermetallic phases. The ordered hexagonal C14 structure develops in approximately 131 defined binary compounds and 263 defined ternary compounds, the ordered f.c.c. C15 structure in approximately 219 binary and 272 ternary compounds, and the ordered hexagonal C36 structure in 17 binary and 14 ternary compounds (Pearson 1985). Therefore, including individual polytypes, over 900 binary and ternary Laves compounds have been defined, with over 360 Laves phases being binary compounds. When compared with the other most common intermetallic phases (Fig. 5), it easy to demonstrate the overwhelming abundance of the material. For example, by just considering binary phases, the Laves phases represent approximately one quarter of the documented phases. By combining the binary and ternary phases, the Laves phases quickly account for the occurrence of intermetallic phases found in nature. The abundance illustrates the attraction for studying the crystallographic nature of the material and alloy design schemes for optimizing the properties of the intermetallic. According to Laves (1967), the abundance of Laves structures among metallic compounds has been attributed to principles governing the geometric arrangement of atoms on ordered lattice sites. For example, metallic compounds tend to form with close packing of atoms, high atomic symmetry, and a metallic nature of bonding, and Laves phases satisfy the geometric conditions more efficiently than other metallic compound structure types.

3. *Geometric Features of Space Filling*

Since Laves phases are predominantly metallic in bonding nature and largely governed by geometric space-filling principles, the metallic atom radii R of the atoms forming the compound have often been used to evaluate atomic size factors involved in the occurrence of these phases. An atom with a metallic diameter D ($D=2R$) will have a different diameter, d, in a Laves phase. Laves phases occur between the metallic radii ratios (R_A/R_B) of $\sim 1.05-1.68$. The atoms forming the compound need to adjust in size to accommodate the ideal space-filling radius ratio ($r_A/r_B \approx 1.225$) in the ordered Laves phase lattice; as a result, the occurrence of Laves phases is more broadly related to the ability of the A and B atoms to expand or contract so that the ideal ratio is approached (Dwight 1961). In fact, space-filling models expressed in terms of the compression of A and B atoms and respective atomic distances have proven to be effective in evaluating the size and volume changes that occur when atoms form a Laves phase (Pearson 1968).

The atomic size accommodation in forming the C15 Laves phase is shown in Fig. 6. The atom size difference between the metallic and intermetallic phases (D–d) yields an apparent contraction (positive value) or expansion (negative value) of an atom

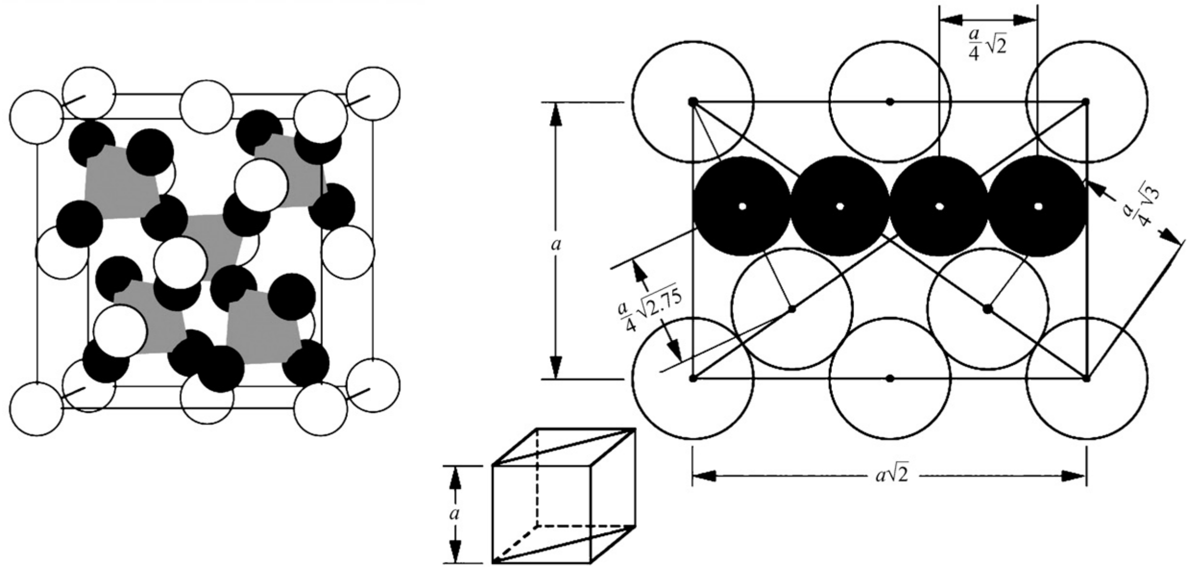

$\frac{a}{4}\sqrt{2}$
$\frac{a}{4}\sqrt{3}$
a
$\frac{a}{4}\sqrt{2.75}$
$a\sqrt{2}$
a

Figure 3
Crystal structure of the C15 Laves phase.

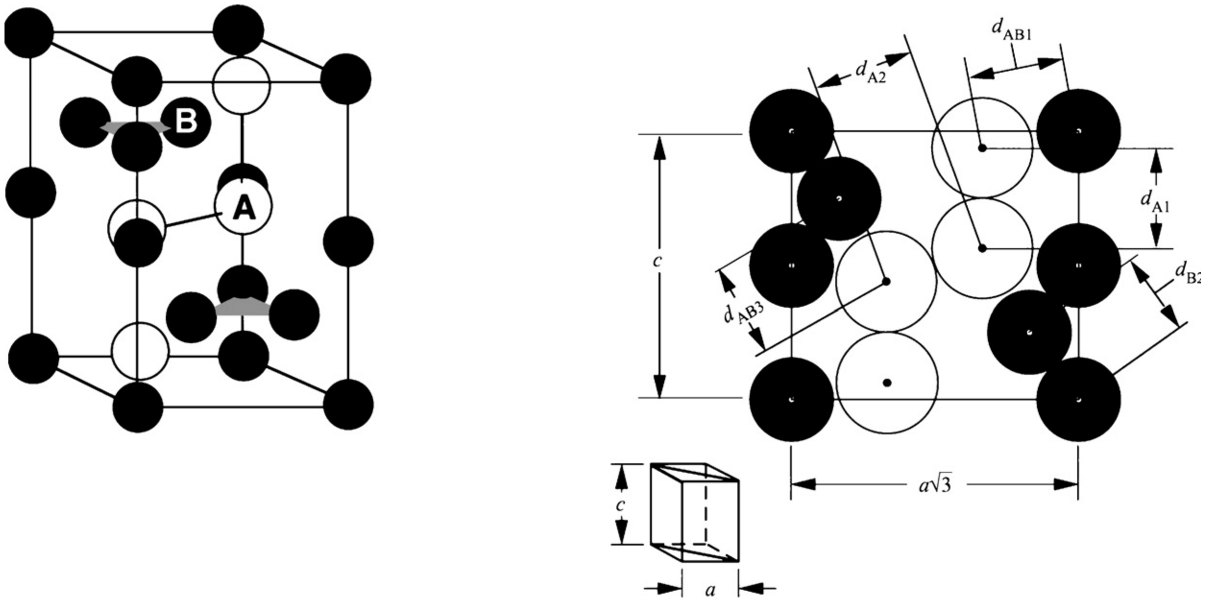

B
A
d_{AB1}
d_{A2}
d_{A1}
c
d_{B2}
d_{AB3}
$a\sqrt{3}$
c
a

Figure 4
Crystal structure of the C14 Laves phase.

within the Laves phase. The bond distance contraction is defined as the atomic size difference between the two structural environments (i.e., D–d). In Fig. 6 the intersections of the regression lines for the A, B, and AB bond distance contractions are illustrated, with only the data shown for the D_{AB}–d_{AB} distances. The intersection of D_A–d_A and D_B–d_B regressions occurs at a value that is slightly lower than the ideal diameter ratio of $\sqrt{(3/2)}$, indicating the potential for compressive bonding forces during intermetallic formation. Although Fig. 6 demonstrates size accommodation for the C15 Laves phase, identical plots can

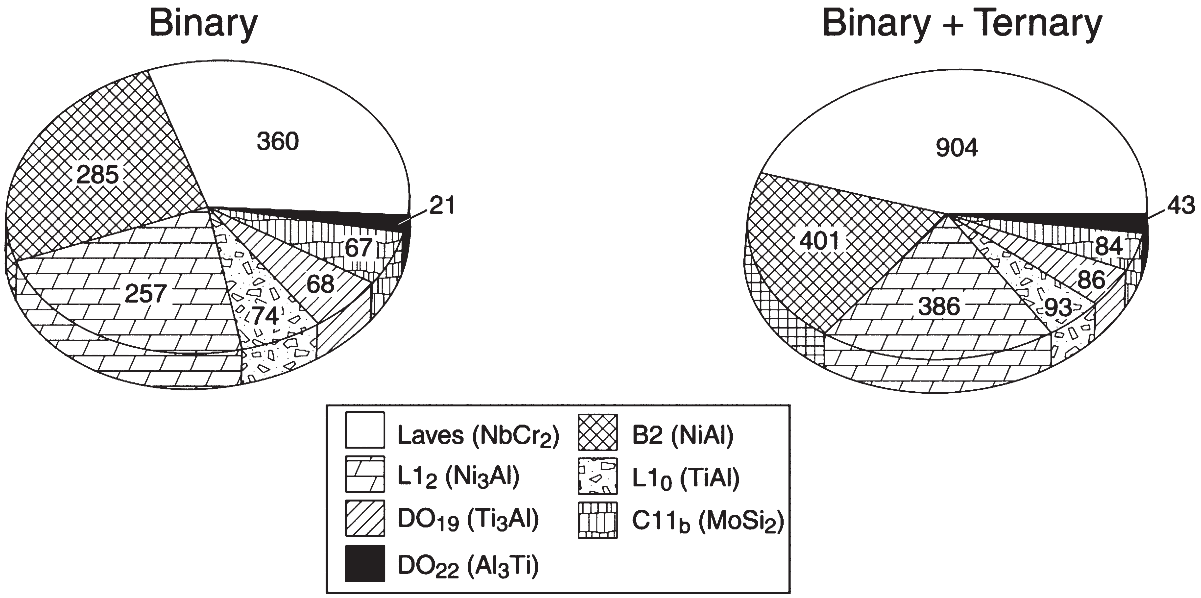


Figure 5
Abundance of the most common intermetallic structure types. The relative abundance of Laves phases are clearly discernible compared with other structures.

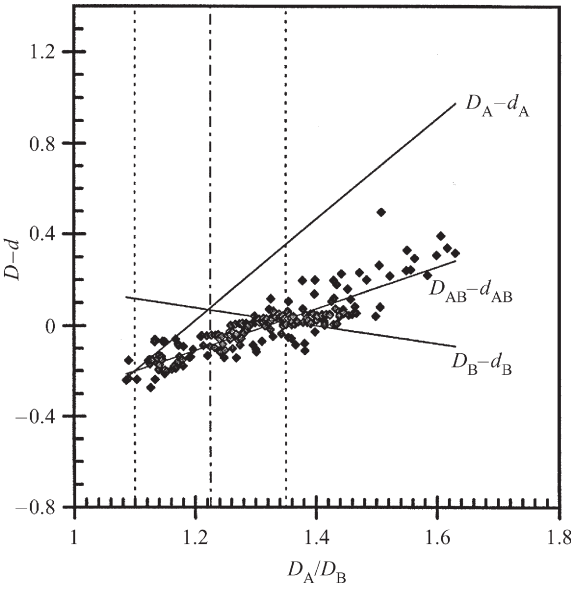


Figure 6
Regressions of the bond contractions in forming the C15 Laves phases from the metallic elements. Data is shown for A–B distances.

be constructed for the other polytypes. As will be seen later, the significance of this plot becomes apparent in the interpretation of the general properties of Laves phases.

4. *Electronic Structure*

Although the geometrical considerations provide insight into the occurrence of Laves phases and their abundance, the interatomic bonding and the polytype stability are more specifically correlated to the electronic structure. The stability of the individual Laves phase polytypes is a function of changing electron concentration. For example, in magnesium-based Laves phases the C15 phase is stabilized at lower electron concentrations ($e/a < 1.8$), and the C14 phase is stabilized at higher electron concentrations (Laves and Witt 1936). In transition-element Laves phases, the C15 phase has been shown to become stable at very high electron concentrations ($e/a > 2.3$) (Elliot and Rostoker 1958), and the C14 phase stabilized at very low electron concentrations ($e/a < 0.73$) (Wallace and Craig 1966). One of the most interesting effects of electron concentration on stability is found in some AB_2 Laves phases where B = Ni or Co (Bardos *et al.* 1961). Many A elements (A = Ti, V, Nb, Ta, Mo, W) do not form a Laves phase with nickel and cobalt, even though the R_A/R_B ratios are favorable for Laves phase formation. However, by substituting silicon on the B lattice sites, ternary C14 compounds can be produced. This phenomenon has been attributed to the binary phases having electron concentrations that are too large. Ternary silicon additions apparently decrease the effective electron concentration, thereby stabilizing the C14 phase.

To quantify the importance of the electronic structure Ohta and Pettifor (1990) used a simple bond order tight-binding analysis to study the relative phase stability of AB_2 structures such as C15, C14, C36 and $C11_b$. Their research indicates that electron densities and band filling are also important in determining the relative stability of these compounds, and some trends can be identified by this simple method.

The stability of transition-metal Laves phases has been examined by using both a linear-muffin-tin-orbital method with an atomic sphere approximation (LMTO–ASA) and a full-potential LMTO method (see Ormeci *et al.* 1996, Chu *et al.* 1998). The relative stabilities of the different stacking variants are energetically very similar, and it appears to be possible to stabilize different polytypes with impurities, constitutional defects, or thermal quenching. In fact, for many Laves phases containing chromium as the B element (e.g., $TiCr_2$, $HfCr_2$, $ZrCr_2$, $NbCr_2$, etc.), excess chromium compositions or quenching from the high-temperature equilibrium phase field are found to stabilize the hexagonal polytypes. A similar situation is found for rare earth elements with cobalt (MCo_2, where M is the rare earth element (Allen *et al.* 1977)).

5. *Constitutional Defects in Laves Phases*

Less than 25% of the binary Laves phases with R_A/R_B ratios of between 1.05 and 1.68 have appreciable ranges of homogeneity (Thoma and Perepezko 1995). The frequency in the number of the intermetallic phases exhibiting any solubility range is increased by a factor of approximately two to three within the specific R_A/R_B ratio ranges of 1.12–1.26 (C14 and C36 phases) and 1.1–1.35 (C15 phases). The upper and lower bounds for the C15 structures can be physically defined as the limits at which the A–B atom distance contractions are greater than the A–A atom distance and B–B atom distance contractions, respectively (see Fig. 6). In addition, the percent contraction, S, of the A atoms and B atoms forming the AB2 Laves phases can be described as $S_X = 100(D_X - d_X)/D_X$ where X = A or B. For all three main polytypes the occurrence of solubility corresponds to a lattice adjusted contraction ($S_{lat} = S_A + 2S_B$) between 0 and 15%. The contraction size rule is a geometric argument based upon the contraction of the atoms forming the intermetallic structure.

Nonstoichiometry in Laves phases is typically characterized by antisite substitutions. A variety of studies has explored the point defect mechanisms, and the prevailing indications are that a direct atom site occupancy substitution occurs. A few examples that exhibit a range of stoichiometry include $ZrFe_2$, Yal_2, $ZrCr_2$, $NbCo_2$, $TiCr_2$, and $NbCr_2$, and the defect mechanisms in these alloys all suggest antisite substitution as the dominant point defect mechanism (Zhu *et al.* 1999). From a hard-sphere model, it is conceptually straightforward to imagine a smaller B atom substituting on to a larger A-site. However, for the few phases (with lower R_A/R_B ratios) that exhibit solubility on the A-rich side of stoichiometry (e.g., $NbCr_2$), the influence of the electronic structure may provide the ability to accommodate the larger A atoms on B sites. For example, in $NbCr_2$, the first-principles calculations indicate that the position of the Fermi level in the density of states suggests that antibonding states are occupied in the C15 structure (Ormeci *et al.* 1996). Since this is not an energetically favorable condition, it is possible that alloying methodologies which lower the valence electron contributions should be favored. Niobium has a lower valence state than chromium and therefore, from an electronic structure argument, niobium occupancy of chromium atom sites may reduce the number of filled antibonding states. Accordingly, the Fermi level will be located at a lower density of states. The more energetically favorable electronic structure may account for solubility on the niobium-rich side of stoichiometry, despite the fact that the niobium atom is larger than the chromium atom ($D_{Nb}/D_{Cr} = 1.15$). For ternary alloying the point defect behavior is apparently similar to the binary systems, and Thoma *et al.* have explored this possibility (1997).

Finally, upon quenching of some Laves phase (including $NbCr_2$), the lattice parameter increases. Most intermetallics behave in the opposite manner, with the lattice parameter collapsing around the vacancies. The origin of this effect has not been determined.

6. *Deformation in Laves Phases*

Studies on the mechanical properties of Laves phases suggest that, as a class of materials, they exhibit a ductile–brittle transition at $0.60T_m$. The binary intermetallic phase is very strong and brittle at ambient temperatures. The Vickers hardness of binary phases typically varies between 600 and 1000 VHN, with toughness values estimated to be less than 5 MPa $m^{1/2}$. Slip occurs predominantly on the $(111)_{cub}$ or $(0001)_{hex}$ planes, with prism slip also being reported.

Due to the specific arrangement of atoms, deformation involving simple planar slip is unlikely and a "synchroshear" or "zonal slip" mechanism involving cooperative movement between planes of atoms may be required (Allen *et al.* 1977, Livingston 1992, Chu and Pope 1993). The mechanism of synchroshear is graphically illustrated in Fig. 7 (extended from the general stacking sequences illustrated in Fig. 2). The deformable layer in the structure has been proposed to consist of the $\alpha c\beta$-type sandwich. The smaller B atoms shear by 1/6[–211] while synchronously, the larger A atoms shear by 1/6[–1–12]. In other words, the smaller atoms move from the illustrated c-site to the b-site while the larger atoms move from the β-site to the γ-site. The synchroshear process has been used to interpret slip, twinning and stacking faults, and phase transformations in Laves phases.

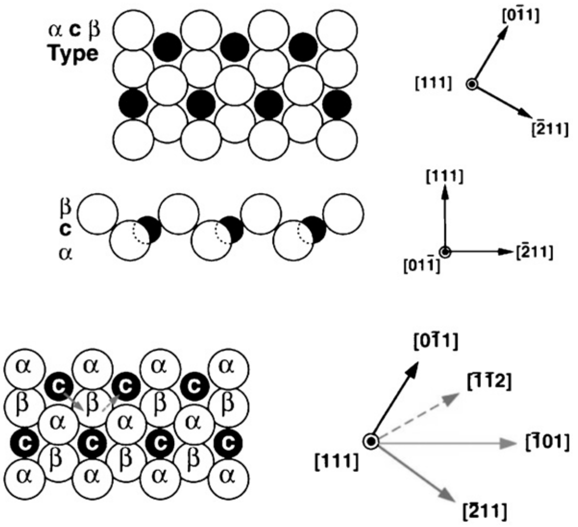


Figure 7
The stacking sequence of the deformable $\alpha c\beta$ "sandwich" in a Laves phase. For synchroshear, the smaller atoms move from the c-site to the b-site while the larger atoms move from the β-site to the γ-site.

Efforts to enhance the complex deformation process have been made using constitutional defects. Ternary additions on A and/or B lattice sites in the AB2 Laves phase may permit enhanced dislocation mobility (Livingston *et al.* 1989, Chen *et al.* 1998). In addition, the existence of vacancy defects may assist the synchroshear process in Laves phases (Hazzledine 1994).

7. Elastic Properties

The known elastic properties of Laves phases are limited, but some are listed in Table 1 and compared to common high-temperature systems. The high Poisson ratios are an indication of a lack of strong directional bonding compared with other intermetallics with Poisson ratios near 0.2. Moreover, the low shear modulus values suggest a potential for resistance to brittle failure. For example, the Rice–Thomson criteria suggest that when $G\mathbf{b}/\gamma < 10$, where G is the shear modulus, $\mathbf{b}$ is the burgers vector, and γ is the surface energy, there is a tendency for resistance to brittle failure. For example, in HfV$_2$ $G = 30\,\text{GPa}$ and γ is estimated to be $2500\,\text{mJm}^{-2}$. If $1/2\langle 110\rangle$ perfect dislocations are mobile $G\mathbf{b}/\gamma \sim 6.3$, and if $1/6\langle 112\rangle$ synchro-Shockley partial dislocation are mobile $G\mathbf{b}/\gamma \sim 3.7$. Likewise, for NbCr$_2$, $G = 80\,\text{GPa}$ and γ is estimated to be $3800\,\text{mJm}^{-2}$. For mobile $1/2\langle 110\rangle$ perfect dislocations $G\mathbf{b}/\gamma \sim 10.4$, and for mobile $1/6\langle 112\rangle$ synchroshear Shockley partial dislocations $G\mathbf{b}/\gamma \sim 6.0$. In both cases, the synchroshear Shockley partial dislocations should provide optimized ductility in the material if they are mobile.

8. Applications of Laves Phases

8.1 Structural Applications

Even though the geometrically structured Laves phases are the most abundant among the various intermetallics, they are the least utilized. Historically, Laves phases have been perceived as a "pest" to the steel and superalloy industries due to grain-boundary embrittlement problems associated with Laves phase precipitates. Stoichiometric binary Laves phases typically display low toughness at low temperatures, contributing to the grain-boundary embrittlement. For this reason, Laves phases were typically avoided, until it was discovered that Laves-phase precipitates (TaFe$_2$) distributed in the matrix of ferritic steels (as opposed to grain boundaries) yielded remarkable wear-resistant properties (as well as good elevated temperature strengths) without sacrificing the low-temperature ductility (Bhandarkar *et al.* 1976). As a result, it has been realized that Laves phases have unique properties that make them attractive for high-temperature structural applications (see, for example, Sauthoff 1989). For example, they retain their high strength (>0.85 of ambient yield strength) at half of the homologous melting temperature, which is the highest of all intermetallics (Fleischer 1989). Many of these phases also have high melting temperatures, excellent creep properties (Anton and Shah 1991), reasonably low densities, and potentially good oxidation resistance for alloys containing chromium, aluminum, silicon, or beryllium.

Dual-phase structures of a Laves phase and a b.c.c. phase are attractive candidates for high-temperature

Table 1
Polycrystalline elastic properties of C15 Laves phases as compared to some common intermetallics.

Sample	Crystal structure	*G* (GPa)	*E* (GPa)	*B* (GPa)	ν	T_m (°C)
HfV_2	C15	32	79	110	0.39	1550
$NbCr_2$	C15	80	215	229	0.34	1730
$HfCo_2$	C15	78	209	229	0.34	1670
$TaCr_2$	C15	88	232	220	0.34	2020
$ZrCr_2$	C15	65	204	157	0.32	1677
$HfCr_2$	C15	73	188	153	0.30	1825
$TiCr_2$	C15	79	204	165	0.30	(1370)
TiAl	$L1_0$	74	181	110	0.23	(1463)
Ti_3Al	$D0_{19}$	57	146	106	0.27	(1164)
Ni_3Al	$L1_2$	92	238	190	0.29	(1360)
NiAl	B2	70	183	156	0.30	1638
$NbSi_2$	C40	153	363	192	0.18	1940
$TaSi_2$	C40	151	359	192	0.19	2040
$MoSi_2$	$C11_b$	186	431	213	0.16	2020
Mo_5Si_3	$D8_m$	127	325	243	0.28	2180
Mo_5SiB_2	$D8_l$	143	361	254	0.26	(~2200)

Source: unpublished data from F. Chu and D. J. Thoma.
Parentheses indicate that the intermetallic does not melt congruent, and the given temperature is a stability limit.

structural applications. Engineering materials can be fabricated using enhanced alloy designs that limit the brittle tendencies of the monolithic Laves phases. For example, $HfV_2 + Nb$ multilayers for superconductivity applications have been successfully cold-rolled to strains up to 30%, with no fracture of the Laves phase (Inoue and Tachikawa 1977). In addition, studies of $NbCr_2 +$ (Cr, Nb, or Ti) have demonstrated marked improvements in ductility and toughness (Takeyama and Liu 1991, Davidson *et al.* 1994). Niobium and chromium both form eutectic structures with $NbCr_2$ and, therefore, this "*in situ* composite" has intrinsic stability to complement the enhanced deformability. Precipitation and coarsening of the Laves phase in the chromium–niobium and chromium–hafnium systems have been found to be extremely sluggish, even at 1200 °C, thus indicating a slow diffusion process in these dual-phase alloys (Thoma and Perepezko 1992, Kumar and Miracle 1994). Ductility and toughness are found to increase with decreasing Laves phase content, and low-temperature toughness values up to 20 MPa $m^{1/2}$ have been achieved (Davidson *et al.* 1994).

8.2 *Nonstructural Applications*

In addition to structural applications, Laves phases also have significant nonstructural applications. In fact, commercial markets have been found for unique properties associated with Laves phases as summarized in a review by Livingston (1992). For example, (Tb,Dy)Fe_2, or Terfenol-D, has been used as a magnetoelastic transducer because of its giant magnetorestriction. Compounds of (Hf,Zr)V_2 have demonstrated superconducting properties with critical fields over 200 kOe and a superconducting critical temperature over 10 k. Laves phases such as Zr(Cr,Fe)$_2$ have been considered for hydrogen-storage applications due to favorable hydriding–dehydriding kinetics and high hydrogen-absorbing capacities. Finally, the Mo(Co,Si)$_2$ Laves phase contributes to the wear-resistance of "Tribaloy" materials.

A large quantity of literature has focused upon Laves phases for hydrogen-storage applications (see encompassing papers by Shaltiel *et al.* 1977, Westlake 1983). Hydrogen-storage intermetallics contain interstices with a suitable binding energy for hydrogen, which allows its absorption or desorption near room temperature and atmospheric pressure. An outstanding feature of Laves phases is their large number of interstices—for example, 17 per AB_2 formula unit for the C15 structure. The absorbed hydrogen may occupy one or more of three types of tetrahedral interstices. If each of these interstices were occupied by one hydrogen atom (i.e., AB_2H_{17}), each host atom would, on average, store 5.7 hydrogen atoms.

Due to electrostatic forces and swelling of the tetrahedra during hydride formation, the maximum hydrogen sorption is speculated to be six hydrogen atoms per AB_2 formula unit (i.e., AB_2H_6). However, only one hydrogen atom per host atom is observed (i.e., AB_2H_3) for hydrides with lattice constants less than 8 Å. The reasons for this limitation need to be understood (on an atomic scale), in order to improve the hydrogen/metal ratio. As a comparison, the most common industrial interest for a hydrogen-storage intermetallic is $LaNi_5$. The maximum hydrogen sorption for this phase is also six hydrogen atoms per formula unit (i.e., $LaNi_5H_6$), but the hydrogen/metal ratio (6/6) is less than the potential maximum for the Laves phases (6/3). In fact, the $LaNi_5$ structure is a "sister" structure of the Laves phase (with an extra plane of B atoms) (Dwight 1961), and many alloying additions that increase the hydrogenation properties of $LaNi_5$ result in a stabilization of a Laves phase.

One interesting component of geometric contributions of hydrogen in Laves phases is in the area of bulk amorphization (Aoki and Masumoto 1993). For example, it has been shown in numerous publications that Laves phases with size ratios greater than 1.35 will amorphize upon the introduction of hydrogen to the lattice. Although significant effort has been directed towards this observation, no model can theoretically describe the event. Curiously, this phenomenon is directly correlated to geometric principles (see Fig. 6), and most likely to electronic structure effects.

Without hydrogen, one of the most successful bulk amorphous alloys has the composition $Zr_{41.2}Ti_{13.8}Cu_{12.5}Ni_{10}Be_{22.5}$ (Schneider *et al.* 1998). In this alloy the formation of the bulk amorphous alloy has been attributed to the "confusion principle." This argument states that with enough atoms of different size, the ordered arrangement of atoms upon cooling from the liquid becomes difficult. This principle is also consistent with the geometric arguments associated with the formation of Laves phases. The Laves phases are stable within size ratios of 1.05 to 1.68. Moreover, with the somewhat complex ordered arrangement of atoms in the topologically close-packed structure, diffusion is typically sluggish (giving rise to the outstanding high-temperature strength). With the kinetic limitations in the structure, competition for nucleation between competing Laves phases (for example, in a deep eutectic trough) yields an ease in amorphous alloy formation. Therefore, the "confusion principle" can also be coupled to nucleation and growth competition between the kinetically sluggish Laves phases, and, in fact, decomposition of the zirconium-based alloy yields multiple Laves phases.

9. Summary

Overall, Laves phases offer attractive features as a class of intermetallics. Their abundance offers the largest available platform for evaluating and designing property responses in materials. Although structural applications for these phases are limited by their brittle nature, dual-phase alloys have shown exciting promise. Moreover, nonstructural applications are diverse, and a wide range of applications have found growing commercial markets.

Bibliography

Allen C W, Liao K C, Miller A E 1977 Fault structures in rare earth–cobalt intermetallics. *J. Less-Common Met.* **52**, 109–15

Anton D L, Shah D M 1991 High-temperature properties of refractory intermetallics. *MRS Symp. Proc.* **213**, 733–8

Aoki K, Masumoto T 1993 Solid state amorphization of intermetallic compounds by hydrogenation. *J. Alloys Compounds* **194**, 251–61

Bardos D I, Gupta K P, Beck P A 1961 Ternary Laves phases with transition elements and silicon. *Trans. AIME* **221**, 1087–8

Barrett C, Massalski T B 1980 The structure of metals and alloys. In *The Structure of Metals*, 3rd edn. Pergamon, Oxford, pp. 256–9

Berry R L, Raynor G V 1953 The crystal chemistry of the Laves phases. *Acta Crystallogr.* **6**, 178–86

Bhandarkar M D, Bhat M S, Parker E R, Zackay V F 1976 Creep and fracture of a Laves phase-strengthened ferritic alloy. *Metall. Trans.* **7A**, 753–60

Chen K C, Allen S M, Livingston J D 1998 Factors affecting the room temperature mechanical properties of $TiCr_2$-base Laves phase alloys. *Mater. Sci. Eng. A* **242**, 152–73

Chu F, Pope D P 1993 Deformation twinning in intermetallic compounds—the dilemma of shears vs. shuffles. *Mater. Sci. Eng. A* **170**, 39–47

Chu F, Thoma D J, Mitchell T E, Lin C L, Sob M 1998 Phase stability of C15 MV_2 (M = Zr, Ta, or Hf): an electronic structure investigation. *Philos. Mag. B* **77**, 121–36

Davidson D L, Chan K S, Anton D L 1994 *AFOSR Report F49620-92-C-0022*. Air Force Office of Scientific Research

Dwight A E 1961 Factors controlling the occurrence of Laves phases and AB_5 compounds among transition elements. *Trans. ASM* **53**, 479–99

Elliot R P, Rostoker W 1958 The occurrence of Laves-type phases among transition elements. *Trans. ASM* **50**, 617–33

Fleisher R L 1989 Mechanical properties of diverse high-temperature compounds—thermal variation of microhardness and crack formation. *MRS Symp. Proc.* **133**, 305–10

Friauf J B 1927 The crystal structure of two intermetallic compounds. *J. Am. Chem. Soc.* **49**, 3107–14

Friauf J B 1927 The crystal structure of magnesium di-zincide. *Phys. Rev.* **29**, 34–40

Hazzledine P M 1994 Twinning by synchroshear in the cubic Laves phase. In: Yoo M H, Wuttig M (eds.) *Twinning in Advanced Materials.* TMS, Warrendale, PA, pp. 403–14

Inoue K, Tachikawa K 1977 On the superconducting properties of the V_2Hf-base Laves phase compound tapes. *IEEE Trans. Magn.* **MAG-13**, 840–3

Komura Y 1962 Stacking faults and two new modifications of the Laves phase in Mg–Cu–Al. *Acta Crystallogr.* **15**, 770–8

Komura Y, Kitano Y 1977 Long-period stacking variants and their electron-concentration dependence in the Mg–base Friauf–Laves phases. *Acta Crystallogr. B* **33**, 2496–501

Kumar K S, Miracle D B 1994 Microstructural evolution and mechanical properties of a Cr–Cr_2Hf alloy. *Intermetallics* **2**, 257–74

Laves F 1956 Crystal structure and atomic size. In: *Theory of Alloy Phases*. ASM, Cleveland, OH, pp. 124–98

Laves F 1967 Space limitations on the geometry of the crystal structures of metals and intermetallic compounds. In: Rudman P S, Stringer J, Jaffee R I (eds.) *Phase Stability in Metals and Alloys*. McGraw-Hill, New York, pp. 85–99

Laves F, Witte H 1936 *Metallwirtschaft* **15**, 840–2

Livingston J D 1992 Laves phase superalloys? *Phys. Status Solidi. A* **131**, 415

Livingston J D, Hall E L, Koch E F 1989 Deformation and defect in Laves phases. *MRS Symp. Proc.* **133**, 243–8

Ohta Y, Pettifor D G 1990 Size versus electronic factors in transition metal Laves phase stability. *J. Phys.-Condensed Matter*. **2**, 8189–94

Ormeci A H, Chu F, Wills J M, Mitchell T E, Albers R C, Thoma D J, Chen S P 1996 A total-energy study of electronic structure and mechanical behavior of C15 Laves Phase Compounds: $NbCr_2$ and HfV_2. *Phys. Rev. B* **54**, 12753–62

Pearson W B 1968 The geometrical factor in the crystal chemistry of metals: near-neighbor diagrams. *Acta Crystallogr. B* **24**, 1415–22

American Society for Materials 1985 *Pearson's Handbook of Crystallographic Data for Intermetallic Phases*. ASM, Materials Park, OH

Sauthoff G 1989 Intermetallic phases: materials developments and prospects. *Z. Metallkd.* **80**, 337–44

Schneider S, Thiyagarajan P, Geyer U, Johnson W L 1998 SANS of bulk metallic ZrTiCuNiBe glasses. *Physica B* **241–243**, 918–20

Shaltiel D, Jacob I, Davidov D 1977 Hydrogen absorption and desorption properties of AB_2 Laves-phase pseudobinary compounds. *J. Less-Common Met*. **53**, 117–31

Takeyama M, Liu C T 1991 Microstructure and mechanical properties of Laves phase alloys based on Cr_2Nb. *Mater. Sci. Eng. A* **132**, 61–6

Thoma D J, Chu F, Peralta P, Kotula P G, Chen K C, Mitchell T E 1997 Elastic and mechanical properties of C15 Laves phase $Nb(Cr,V)_2$. *J. Mater. Sci. Eng. A* **240**, 251

Thoma D J, Perepezko J H 1992 An experimental evaluation of the phase relationships and solubilities in the Nb–Cr system. *Mater. Sci. Eng. A* **156**, 97–108

Thoma D J, Perepezko J H 1995 A geometric analysis of solubility ranges in Laves phases. *J. Alloys Compounds* **224**, 330–41

Villars P, Calvert L D (eds.) 1985 In: *Pearson's Handbook of Crystallographic Data for Intermetallic Phases* 1st edn. ASM, Materials Park, OH

Wallace W E, Craig R S 1966 Electron concentration and phase stability in Laves phases and Ni–Al alloys. In: *Phase Stability of Metals and Alloys*. ASM, Cleveland, OH, pp. 255–72

Westlake D G 1983 Site occupancies and stoichiometries in hydrides of intermetallic compounds: geometric considerations. *J. Less-Common Met*. **90**, 251–73

Zhu J H, Pike L M, Liu C T, Liaw P K 1999 Point defects in binary Laves phase alloys. *Acta Mater*. **47**, 2003–18

D. L. Thoma
Los Alamos National Laboratory, New Mexico, USA

This Page Intentionally Left Blank

L

Liquid Crystals: Overview

A totally ordered system is rigid and unresponsive; a disordered system represents chaos. Liquid crystals succeed in being intermediate between these extremes. As responsive, self-assembled systems combining order with fluidity, they are intellectually fascinating and technologically valuable. This brief overview seeks to provide a perspective of them and stimulation for the reader to learn about them.

1. Introduction

Liquid crystals represent a state of matter additional to the three states of matter we all know, i.e., the gaseous, the liquid, and of course the solid states, including both crystalline and amorphous solids. In the gas phase, the constituent units (atoms or molecules) are widely separated, have high kinetic energy, and move about freely and rapidly. In the liquid phase, the atoms or molecules are much closer, but in this condensed state they are still free to translate and rotate, and have no overall organization. In the solid, this freedom of motion is lost, and the units are locked in position, in the case of the crystal in a regular, three-dimensional repeating lattice.

How then does the liquid crystal (LC) state fit into this set of well-accepted states of matter? It does so by lying intermediate between the crystalline solid and the liquid states. Generally speaking, when a crystal melts, it does so at a well-defined temperature and passes to the liquid state through a transition which is reversible, often with some degree of supercooling. With a liquid crystal material however, the organized molecular crystal again melts at a well-defined temperature, but gives an intermediate LC phase or a sequence of such phases before the molecularly disorganized isotropic liquid state is produced, again at a well-defined transition temperature.

Here we must note that the ordered structure of a crystal can also be broken down by solvents such as water during the process of dissolution, now giving the molecularly disorganized state of a solution. With suitable materials such as soaps and other amphiphiles, and controlled amounts of solvent, an intermediate liquid crystal state or states having fluidity to some degree and ordering of micelles made up of individual molecules may occur. Such LC phases, consisting of amphiphile and solvent, are described as lyotropic (solvent-induced) LC phases, so distinguishing them from those produced from pure material or mixtures of pure materials by the action of heat, these being called thermotropic LC phases.

Occupying a position intermediate between the crystalline solid and the isotropic liquid and combining fluidity with organization, LCs not surprisingly have properties intermediate between their contiguous crystalline and disordered states. The intermediate LC states, of which there are several types, are also known as mesophases (Greek: *mesos*—between and *phasis*—state).

Because of this combination of fluidity and order, LCs have come to be a most challenging area of research in a wide range of scientific disciplines as disparate as chemistry, physics, biology, engineering, electronics, mathematics, and theory. As such, LCs represent one of the most multidisciplinary areas of science, and in bringing together the activities of scientists from many disciplines, LCs have yielded a raft of new knowledge, new materials, new theory, and new technological developments. They quite simply are fascinating systems.

2. Discovery of Liquid Crystals

In the period 1834–1861 observations were made that were later to be recognized as relating to LCs, but at the time there was no follow through to an understanding of their significance. The observations were made on biological samples derived from nerve tissue, for example, myelin, a complex lipoprotein forming a sheath round nerve cells. In aqueous sodium oleate, these sheaths gave myelinic forms visible in the polarizing optical microscope; they were fluid and birefringent (anisotropic fluids), and therefore structurally ordered. The implications from such observations were not however evident in the mid-nineteenth century.

The discovery of LCs is therefore generally attributed to Friederich Reinitzer of the German University of Prague, because his studies of thermotropic systems were followed through to acceptance that LCs represent a true state of matter intermediate between, but differing from crystals and isotropic liquids. Reinitzer's research paper (Reinitzer 1888) was presented on May 3, 1888 and set out his observations on the pure materials cholesteryl acetate and cholesteryl benzoate and the colored reflectance phenomena exhibited by their melts. Most importantly, for cholesteryl benzoate, he registered its double melting behavior; the crystals changed at 145 °C to give a cloudy fluid which suddenly clarified at the much higher temperature of 178.5 °C.

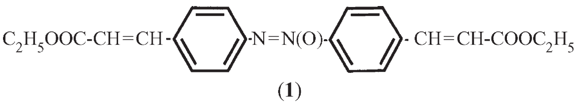
The two transitions were reversible on cooling, the transition from the cloudy state to the crystalline state occurring with supercooling. Reinitzer also studied thin films of the melt of cholesteryl benzoate and observed the range of spectral colors reflected by the fluid state until it crystallized, and that when such films were viewed in transmission, the colored light was complementary to that reflected at the same temperature. Reinitzer wrote to Professor Otto Lehmann at the Polytechnical School at Aachen about his observations and sent him two samples. Details of the interaction have been well documented by Kelker and Knoll (1989), but the main point was agreement that Reinitzer's materials were pure and that the phases formed were homogeneous. A year later, Lehmann, then at the University of Karlsruhe, published his paper entitled *Uber fliessende Kristalle*, using for the first time the term flowing crystal or liquid crystal.

In his own paper of 1888, Reinitzer acknowledges that others before him had seen color effects in molten cholesteryl systems, but these observations were very much less detailed; only Reinitzer had observed the double melting phenomenon and the complementary relation between the transmitted and reflected light. Today we accept that the color effects are characteristic of the helical structures of many cholesteric LC phases, or as they are now known, chiral nematic (N^*) LC phases. Through Reinitzer, a true understanding of LCs began, and he must be acknowledged as the true discoverer of LCs, the appropriate date being that of his letter to Lehmann, March 14, 1888.

3. Developments in the Late Nineteenth and Early Twentieth Centuries

3.1 Materials

LCs occur in synthetic as well as naturally occurring materials like cholesteryl esters. Synthetic azoxy ethers were made in 1890, and as the prerequisite for an elongated or rod-like molecular shape became recognized, many other examples of synthetic LC materials were found. Examples were 4-methoxycinnamic acid, an elongated hydrogen-bonded dimer and ethyl 4-azoxycinnamate (**1**) with the structure:



(**1**)

This material demonstrated that all LC phases are not the same. Unlike the nematic (N) phase of 4-methoxycinnamic acid and the chiral N^* phase of cholesteryl benzoate, compound (**1**) gave a different type of LC phase later characterized as a smectic (Sm) LC phase. The existence of different LC phases was clearly shown by Lehmann in 1907 with the material cholesteryl decanoate (**2**), for which the phase sequences on heating and cooling, respectively, were:

$$\text{Cr } 82.2\,^\circ\text{C LC(1)} 90.6\,^\circ\text{C I and I } 90.6\,^\circ\text{C LC(1)}$$
$$77.4\,^\circ\text{C LC(2)} < 70\,^\circ\text{C Cr} \qquad (\mathbf{2})$$

Later, Vorländer showed that ethyl 4-(4-methoxybenzylidene)aminocinnamate (**3**) had the phase sequence:

$$\text{Cr}108\,^\circ\text{CLC(2)}117\,^\circ\text{CLC(1)}136\,^\circ\text{CI} \qquad (\mathbf{3})$$

both on heating and cooling, but with supercooling of the recrystallization process. Here, Cr represents the crystalline solid, I is the isotropic liquid and in today's terminology LC(1) is the nematic phase [N for (**2**) and N* for (**3**)] and LC(2) is the lamellar smectic A (SmA) phase (see Fig. 1).

The examples of compounds (**2**) and (**3**) allow the definition of two further terms. For (**2**), the N (LC1) phase is observed both on heating and cooling cycles, but the SmA (LC2) phase only occurs on supercooling the N phase below the melting point of the crystals. The N phase is called enantiotropic and the SmA phase monotropic. In the case of compound (**3**), both the N (LC1) and SmA (LC2) phases are enantiotropic, being observed on both heating and cooling cycles.

Although the number of LC materials was growing steadily, and in a short period Vorländer and his group contributed around 250, scepticism remained in some quarters that LCs represent a true state of matter. Tammann believed that LCs were colloidal suspensions and Nernst that they were mixtures of tautomers, and bitter arguments arose with those who maintained that LCs were homogeneous states of matter. However, review articles by Schenk and by Vorländer in 1905 and growing physical evidence from Stumpf on the optical properties of LCs led to general acceptance of Lehmann's belief which was firmly established by the work of Georges Friedel (1922). This article provided the foundation for future progress on LCs and deserves to be read by all working in the field.

(a) Nomenclature

Friedel strongly disliked Lehmann's use of the term liquid crystal to describe these anisotropic, fluid states of matter, much preferring to call them mesophases (intermediate phases). His argument was that LCs are neither true crystals nor true liquids, but something totally distinct. Today, however, the terms liquid crystal and mesophase coexist quite happily and are used synonymously. A useful term derived later from these earlier debates is the term mesogen to describe a material that is structurally equipped, for example, rod-like molecules, to form a mesophase or liquid crystal.

(b) Phase morphology

More importantly, Friedel's article gave us the terms to describe the different LC phases that can exist, often in the same compound with reversible transitions defining their limits. These terms described the nematic phase (N) involving a statistically parallel arrangement of rod-like molecules without any ordering of the ends of the molecules, the cholesteric phase (Ch, but now known as the spontaneously twisted or chiral nematic phase, N^*) and the smectic phase. Friedel defined only one type of smectic phase (today's SmA phase—see Fig. 1). Different types of Sm phase are now known, but Friedel's article is elegant for its full analysis of the SmA phase as a system of rod-like molecules lying roughly parallel with their ends in planes and their long axes orthogonal to these planes which are free to move over one another. It was also demonstrated that in an aligned SmA phase mounted as a film between glass surfaces, the layers may lie flat to the surfaces, the sample behaving in the microscope as a uniaxial crystal, as it is being viewed down the molecular long axes.

However, Friedel realized that space may be filled other than by flat layers, and that in unaligned samples, the layers may assume a curvature and fill space with a continuum of focal-conic domains which are composed of a series of curved, parallel surfaces known as Dupin cyclides. He also understood that the ellipses and their related hyperbolas, visible as dark lines in the microscopic textures of such Sm films, are optical discontinuities or defects in the structure, i.e., lines along which the optic axis suddenly changes direction. The focal-conic microscopic textures of smectics are characteristic and provide one of the means of identifying mesophase types through polarizing optical microscopy. With experience, all

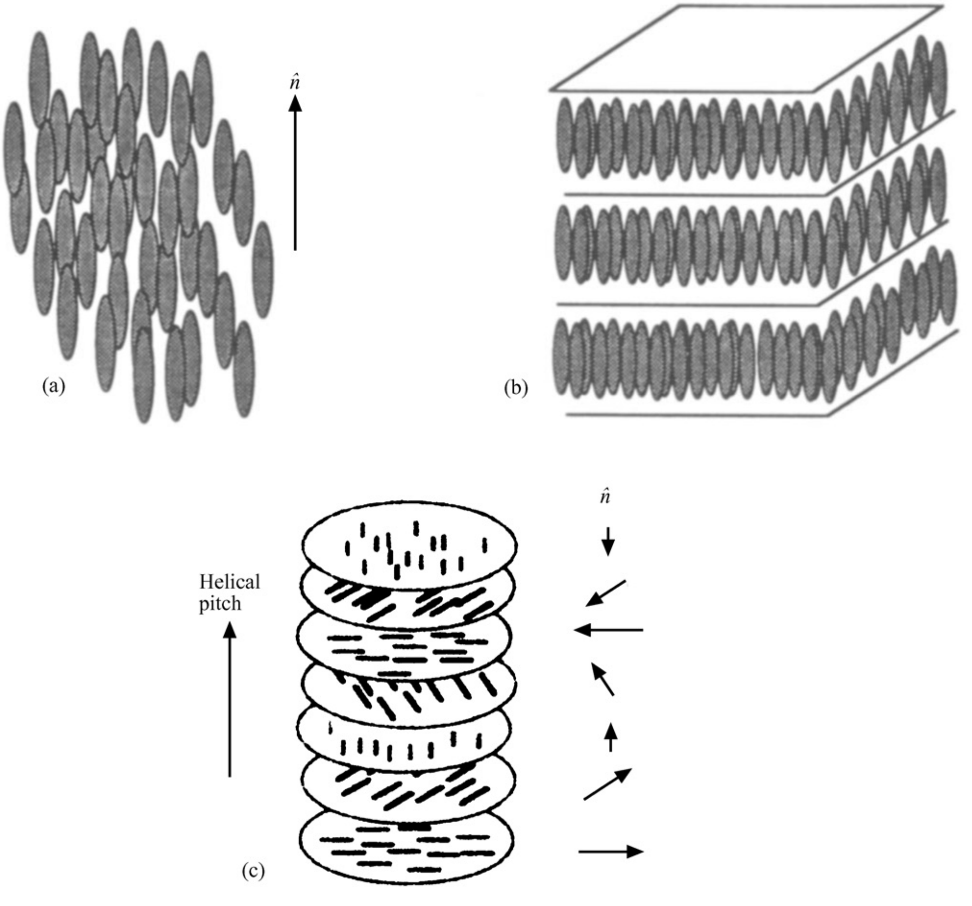


Figure 1
Molecular ordering in the simpler calamitic liquid crystal phases. (a) Nematic phase (N)–statistical parallel alignment of rods; (b) smectic A phase (SmA)–statistical parallel alignment of rods forming diffuse layers to which the axes of the rods are on average orthogonal; (c) chiral nematic or cholesteric phase (N^*)–schematic representation with the rod-like molecules forming sheets in each of which the order is as in a nematic, but on passing through a stack of sheets, the long axes and the director rotate forming a helical structure with a pitch orthogonal to the planes of the sheets.

but the most subtle transitions between different liquid crystal phases can be detected microscopically, and the phase textures can be used to characterize (Gray and Goodby 1984) the different phase types—N, N^*, SmA, SmB, SmC, SmF, SmI.

As will emerge later, the diversity of known LC phases has grown with the passage of time. Much later, the discovery by Chandrasekhar *et al.* (1977) of LC phases formed by disk-shaped molecules added to the number, and has required an addition to the nomenclature to distinguish such phases of discogens from those of rod-like molecules. The latter are described as calamitic LC phases and those of disk-shaped molecules as discotic nematic and columnar LC phases.

Some mesophases formed by rod-like molecules are cubic in nature, and it is interesting that Reinitzer first described such an LC phase. He clearly described the blue color that appears in the isotropic liquid of cholesteryl esters before the sample cools a little further and gives the cloudy N^* phase. The transient blue color is now known to be associated with the optically isotropic blue phases (BP) of which three are now known. Other types of cubic mesophase are now known; the first, not found until 1950, was originally thought to be smectic (SmD), but it is now known not to be a lamellar phase, but a cubic mesophase.

3.2 *The Middle Period*

This spans the period from about 1925–1958. The years up to 1934 are marked by the extensive publications of Vorländer, who enriched the field with so many new materials, sometimes as many as three LC phases being formed by a single material. His work firmly established that the LC phases known at that time required rod-like molecules, but also demonstrated that if the molecular length is great enough, appreciable protrusions from the side of a molecule in the form of lateral substituents can be tolerated. In considering Vorländer's contribution to the subject, the single most important feature was, however, the proof that the structural units of mesophases are molecules. Until then, even Lehmann was unsure about the building units involved and what happens at the phase transitions.

Early understanding of phase structure and phase transitions was enhanced by X-ray studies. In 1923 it was shown that aqueous sodium oleate, a lyotropic LC system, gave X-ray reflections consistent with a lamellar structure, and E Friedel (the son of G Friedel) extended X-ray studies to ethyl 4-azoxybenzoate, confirming G Friedel's conclusion from microscopic studies on defect textures that the LC phase is lamellar (smectic). The first understanding of the different types of smectic phase began in this period too, through Herrmann's X-ray studies of a material exhibiting more than one smectic phase. These showed clearly that a transition occurred from a lamellar phase with a statistical ordering of the rods in the layers at higher temperatures, to a lower temperature smectic with a hexagonal ordering in the layers. Herrmann also found that some thallium soaps had the molecules arranged in layers, but with their long axes tilted with respect to the layer planes.

In this period a very important event was the Faraday Society Meeting held in London in 1933 on the subject of *Liquid Crystals and Anisotropic Fluids*. Twenty four papers were presented at the meeting, the dominance of Vorländer and his school being emphasized by the fact that five of the presentations were his. Other important names on the list of contributors were Fréedericksz and Zolina (forces causing orientation of anisotropic fluids), Zocher (alignment of LC phases by magnetic fields), Sir W H Bragg (developing on Friedel's focal-conic domains and the involvement of Dupin cyclides), Ostwald (viscous properties of LC phases), Lawrence (lyotropic systems), Herrmann (X-ray studies), and Bernal and Crowfoot (on structures of crystalline materials that form LC phases on heating).

Some debate at the meeting was sparked by the paper by Ornstein and Kast on new arguments for the swarm theory of liquid crystals. Swarm theory, first proposed by Bose in papers between 1907 and 1909, laid down that the nematic phase consisted of elongated swarms of some 10^6 statistically parallel aligned molecules, the bulk nematic being a random ensemble of such swarms. In the published *Faraday Discussions*, Zocher expressed strong reservations about swarms, and despite experimentation in the 1930s to prove the existence of swarms, none was successful, and none ever has been. By 1938 Zocher was expressing still stronger reservations about the theory, proposing instead that the nematic phase is a continuum in which the molecules, although lying locally statistically parallel, can change their orientation from one part of the bulk to another, but always in a continuous manner, and that when for one reason or another the change is made discontinuous, then defect structures appear in the sample.

Zocher's distortion hypothesis was the beginning of modern continuum theory of LCs, and the 1933 Faraday Meeting indeed marked the beginning of the end of swarm theory. However, it was not until much later, as a result of the Second World War, at the second Faraday Meeting in 1958, that the famous paper by Frank (1958) was presented as a theory of curvature elasticity. Based on his reexamination of Oseen's earlier theory of liquid crystals, this seminal paper made possible the future developments of continuum theory of LCs by de Gennes and the Orsay Group in France (see de Gennes and Prost 1993) and by Ericksen in the USA and Leslie in the UK—see the review by Carlsson and Leslie (1999).

The highly important textures given by thin films of LCs when viewed between crossed polarizers in the optical microscope are due to the anisotropy of their optical properties—their birefringence. This middle period should also be remembered for the important work by Chatelain in Montpellier on the refractive indices of aligned nematics. His theoretical treatment of their refraction and their ordinary and extraordinary indices of refraction has been highly significant.

In the 1950s, organic chemists were also contributing by conducting systematic studies of new materials aimed not simply at producing new materials, but rather at establishing structure/property relationships. The names of Weygand and Wiegand should be mentioned here as well as that of Gray who published 20 papers in the period 1951–1959 showing that systematic changes in molecular structure cause systematic changes in transition temperatures and phase behavior. The relationships established in this period were to prove of great value in later years in achieving room temperature LCs for applications.

This middle period, sometimes described by the less than well informed as a period in which little happened, also saw some other very important events, for example, the discovery by Robinson and colleagues of lyotropic N^* phases in solutions of the polypeptide poly-γ-benzyl-L-glutamate in organic solvents, and the work of Eaborn and Hartshorne on the mesophase exhibited by di-isobutyl silandiol, a mystery that was unexplained until light was shed on it through the discovery of the mesophases of disk-like molecules 15 years later. Finally, there was the publication of the excellent review by Brown and Shaw (1957), surely a catalyst for the escalation of activity that occurred in the field from the late 1960s and continues today.

3.3 The Period 1958 to the Present Day

(a) The first ten years

These years were not only the prelude to technological applications of LCs, principally in the area of electro-optical displays, but to applications of thermochromic LCs. It is emphasised that these developments in applications, so important in stimulating research on LCs and generating funding, were, certainly in the area of optoelectronics, a direct consequence of knowledge stemming from fundamental research done during these ten years. Three particular points are noted below.

(i) As a result of Frank's reformulation of Oseen's elastic theory of LCs, the combined efforts of Leslie and Ericksen generated today's viscoelastic theory of LCs. Quoting from the review by Carlsson and Leslie (1999), "this theory formulates rheological equations of state for non-linear viscoelastic materials and is needed in formulating the proper dynamical equations for anisotropic materials, such as liquid crystals." The theory is essential to understand properly the dynamics of LCs under, for example, the conditions of their reorientation by external fields, as in the switching processes in an electro-optic display cell. The theory was established in this period; the systems would come later.

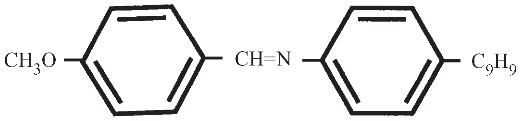
(ii) In 1958–1960, the Maier–Saupe theory of the nematic state was published. This dispelled the idea that dipole moments generate the molecular interactions that form LCs, and focused attention on anisotropic London dispersion forces as being responsible for the molecular interactions. This stressed the importance of the order parameter of the system and focused attention on real molecules.

(iii) The review article already mentioned by Brown and Shaw and five years later Gray's monograph on the *Molecular Structure and Properties of Liquid Crystals* (Gray 1962) stimulated sufficient interest in LCs to encourage Glenn Brown to host an International Liquid Crystal Conference at Kent State University in 1965. This was a small meeting (some 90 delegates), but a significant one, unifying the small body of LC researchers such that the meeting was followed by another in 1968 at Kent and one in 1970 in Berlin, and from then on biennially.

Research at the Westinghouse Corporation in the USA had been going on from the late 1950s on the uses of N* materials as temperature sensors. Some of this work led by Fergason was reported at the 1965 and 1968 conferences, and the awakening interest of industry in LCs was further evident at the 1968 conference by the presence of the group under Heilmeier from the Radio Corporation of America at Princeton with clear interests in useful LCs for applications. The quest for materials giving the N state at room temperature was thereby stimulated.

(b) The 1970s

This was a rich decade for progress in LCs. The awakening commercial interest in the USA was soon followed by the discovery by Kelker and Scheurle of the first room temperature nematic LC. This material was 4-methoxybenzylidene-4′-butylaniline (MBBA) **(4).**

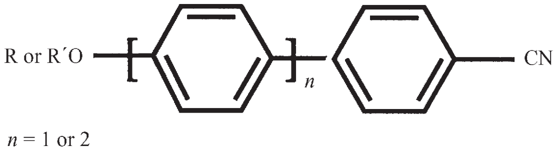


(4)

It melted at 21 °C and stayed nematic until 46 °C, the nematic phase supercooling for long periods at temperatures below 21 °C. The material was not ideal. It was yellow, and the inter-ring linkage made the material sensitive to moisture. However, it did enable experimentation, most significantly in the area of dynamic scattering (DS) displays wherein a clear nematic film could be driven by a field into a light scattering state by turbulence caused by ionic impurities/dopants. The instability of the material made device manufacture difficult, but the devices were marketed and they stimulated much interest and more research.

This resulted in the UK Ministry of Defence giving support to Gray for a materials program to find stable, room temperature LC materials for electro-optic applications, with the emphasis on material stability. The program was not big at first and its eventual success, based upon structure/property relations for LCs established by Gray in the 1950s and 1960s and a swing in interest away from negative to positive dielectric anisotropy materials, was achieved with the assistance of just two colleagues. The swing to positive materials arose because of publications on the twisted nematic (TN) device in publications/patents by Schadt and Helfrich (1971) and Fergason. This device is a field effect device and required high resistivity materials which were desirably colorless and stable.

The new materials were 4-alkyl- and 4-alkoxy-4′-cyanobiphenyls (**5**) (Gray *et al.* 1973), some of which gave N phases in the pure state at room temperature, and as mixtures incorporating the terphenyl analogs ($n = 2$, $R = C_5H_{11}$) to enhance $T_{N\text{-}I}$ covered the nematic range from $-20\,°C$ to $+60\,°C$ and performed excellently in TN displays. Details of this program and its commercialization by BDH Ltd (later Merck Ltd, UK) have been recorded by Hilsum (1991).



(**5**)

$R = C_5H_{11}$, $n = 1$(5CB)Cr22.5 °CN35 °CI

The cyanobiphenyls made available the materials needed for production of reliable LC displays and provided the secure basis for the infant LC display industry later to become the multibillion dollar industry of today. The biphenyls also provided room temperature smectics, and with chiral groups R or R′, room temperature chiral N^* materials. The simple concept of removing the labile central linkage of MBBA and compensating for the molecular shortening by using a terminal cyano group, which also gave a high positive dielectric anisotropy, soon led others to produce analogs, such as the phenylcyclohexanes of Merck and the phenylpyrimidines of Hoffman La-Roche, widening valuably the range of physical parameters (birefringence, etc.) available in room temperature systems.

The market place in the Far East welcomed the first direct drive TN displays for watches and calculators (Europe was not quite ready for portable, battery-driven equipment), and soon demands arose for displays portraying more complex data. Multiplexed devices were needed first, and new mixtures with added esters were developed for these. The growing demand for still more sophistication then led to the supertwisted nematic (STN) display and also active matrix addressing of devices using thin film transistors. The area, including full color displays for direct viewing and projection, and if needed having high definition resolution, has been reviewed by Schadt (1994). Today's LC displays are very beautiful creations, but work goes on to perfect still further their viewing angle, brightness, and definition. These displays are everywhere, in the home, the workplace and the car, and take the industry into the new Millennium with high levels of confidence.

Other LC display forms have, of course, evolved, for example, the ferroelectric (FLC) display stemming from the seminal work of Meyer in the 1970s involving chiral SmC LCs. These displays are capable of microsecond rather than millisecond switching, and although slow to become commercialized because of addressing problems, they will surely command an important place in the LC display industry of the future.

The move towards displays using smectic LCs—including of course the new antiferroelectric LC display—benefited greatly from the earlier fundamental X-ray and neutron scattering studies of Levelut in France and Leadbetter in the UK leading in the 1980s to a clear structural classification of the then known smectics. A distinction was made between true smectics with little or no correlation between layers (SmA, SmB, SmC, SmF, and Sml) and those phases previously regarded as smectics and now treated as soft crystals, in which interlayer correlations really give a loose three-dimensional order. These are simply referred to as B, E, G, H, J, and K phases or CrB, etc. phases. The lettering system derives simply from the order of phase discovery and is due to Sackmann and colleagues in Halle who discovered the LC miscibility rules that can be used as an added aid to phase identification.

(c) New systems

With increased research activity, curiosity driven and mission oriented synthetic work on new LC materials led to the discovery of new and interesting phase phenomena—new phase types, phase sequences, and subphases. These have complicated the classification system of the 1980s, but have not required its radical change. The following examples should be mentioned.

(i) The elegant microscopic studies of Kleman on defect structures and texts by Demus (1978) and Gray and Goodby (1984) on mesophase characterization and identification, stimulated phase studies, such as those of Cladis, on re-entrance phenomena whereby phase sequences on cooling such as I–N–SmA–N_{re} were established. Similarly, blue phases first noted by Reinitzer received much attention, and the phases BPI, BPII, and BPIII were established, the first two having cubic defect structures.

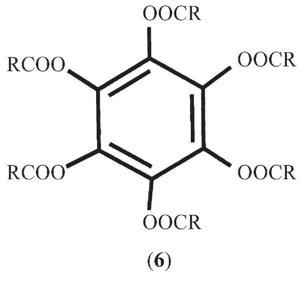
(ii) The quest for new materials for the FLC display led to a host of new SmC materials. Amongst these were many esters in which new phase types have been found. One of these was the alternating tilt equivalent of the uniformly tilted SmC phase, the SmC_{alt} phase, and its chiral analog the SmC^*_A phase with antiferroelectric properties. Ferrielectric phases with tilt direction changes after sequences of like tilt in the layers (SmC^*_{Fl}) were also discovered together with a range of subphases such as SmC^*_α and SmC^*_γ, often identified by differences in their electro-optical behavior. The antiferroelectric phase with its tristable switching characteristic has device potential reported to be superior to that of an FLC display and prototype commercial displays have been produced.

(iii) Twist grain boundary phases provide further examples of new phase discovery. The twist grain boundary SmA* (TGBA*) phase was discovered by Goodby *et al.* (1989). The phase is an interesting example of a frustrated LC structure in which the opposing tendencies of the system to be twisted and lamellar are met by the formation of lamellar blocks between a succession of which there exists a twist. This generates a regular array of screw dislocations which remarkably were predicted by de Gennes 17 years earlier when he realized the analogous roles of the director in a SmA and the magnetic vector potential of a superconductor. Later a tilted version (TGBC*) of the TGBA* phase was found at the N*–SmC* transition.

3.4 Other Major LC Topics

(a) Discotic LC systems

These were first discovered by Chandrasekhar *et al.*, at Bangalore (1977) (see *Discotic Liquid Crystals: Overview*). The systems they studied had typically a central flat core to which long chains were attached via ether or ester linkages. The example shown is hexa-substituted benzene, a hexa-ester.



(**6**)

Such flat disk-like molecules can form nematic phases (N_D) in which the planes of the molecules lie parallel, but in a disordered array with respect to their centers, or columnar phases (Col) of different types where the disk-shaped molecules stack up in columns and the columns pack with their long axes parallel in an ordered or a disordered manner. Many discotic materials are now known, with a variety of cores and side groups. Such systems have now been extended through molecules which are not themselves disk-like, but may aggregate, for example, by hydrogen bonding, to form disk-shaped aggregates which can give N_D and Col phases. Examples are diisobutylsilandiol, several polyols, and some carbohydrate materials, for example, galactopyranoses. Activity in the discotic field has been stimulated by possible applications of doped columnars as nonlinear conducting materials.

(b) Liquid crystal polymers (LCPs)

There are two classes of these moderate (oligomeric) to high molar mass LC systems, main chain LCPs with mesogenic groups linked end to-end, preferably via flexible spacers, to form the backbone and side group LCP (see McArdle 1989) with the mesogenic groups linked by one end to the backbone. The mesogenic units may be calamitic, discotic or amphiphilic (or mixed) and thermotropic or lyotropic mesophases may be formed.

The discovery of the side group type is due to Finkelmann who has also contributed greatly by synthesizing the first LC elastomers, which have greatly extended the scope for technological applications of LCP systems, including those in the areas of electro-optics and optical nonlinearity. The field has widened too to include combined systems, where the polymer has main chain and side group LCP characteristics, and also composites of high and low molar mass LC materials.

Polymer dispersed liquid crystals (PDLC) are related to composites where micron scale droplets of low molar mass nematic LC are dispersed in a glassy isotropic polymer matrix. Flexible sheets may be produced which scatter light because of the randomness of the directors of the N droplets, but application of an external electric field aligns the director and gives a clear film. Such PDLC systems are of interest for switchable panels and indeed electro-optical displays.

(c) Metallomesogens

Mesogens incorporating metals were studied by Vorländer in the early 1900s, but the field of organometallic mesogens really began to develop in the mid- to late 1970s through the work of Billard and of Giroud-Godquin. The material types are now very diverse, as is the range of metals used in them, and examples extend to calamitic, discotic, and polymeric LC systems. The field of metallomesogens has been well reviewed in recent times (Serrano 1996).

4. *The Recent Years*

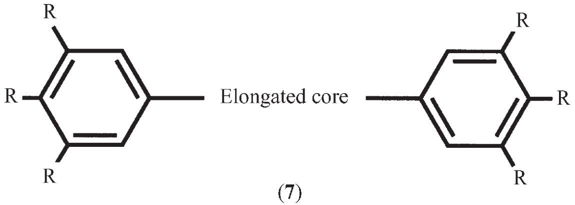
The productivity of the 1970s in generating technological applications and the new areas of discotics, LCPs, and metallomesogens led to concerns that the subject of LCs might be going to fragment into subgroups with separate scientific meetings, so losing the invaluable coherence gained from the biennial International Liquid Crystal Society Conferences. Fortunately this has not occurred, partly because advances in theory, techniques, physical studies, etc. are critical for all areas. However, other factors have contributed to binding the subject together, and it is useful to consider some of them; also, they nicely illustrate some of the new science evolving today.

4.1 Oligomers

These naturally bridge the low and high molar mass LC areas. They were originally studied to establish how polymeric characteristics evolve in low molar mass LCs and have been important in this way, but new aspects too have arisen through studies of LC dimers/trimers. Recently, a new type of smectic packing (intercalated Sm) was found by Luckhurst and his group for certain calamitic symmetric dimers, and in some asymmetric calamitic dimers, Hardouin and his group found that middle order terminal chain lengths give rise to interesting anomalies of the smectic periodicity, connecting with the interposition of N phases between two different Sm types in the homologous series of symmetric dimers.

4.2 Polycatenars

These are molecules in which a rod-like core terminates in one or two moieties that are disk-like. For a hexacatenar (phasmid) such as structure (**7**), the phases are predominantly columnar, but for some tetracatenars, short-chain homologs may be SmC and N, while longer chain homologs are hexagonal Col.



(**7**)

Also, lamellar SmC, cubic and hexagonal Col phases have been found by the group at the Centre de Recherche Paul Pascal in France on heating a single pure polycatenar, and Weissflog at the Max Planck Gesellschaft in Halle found the sequence lamellar SmC, oblique Col, lamellar SmC, N, I on heating a six-ring double-swallow-tailed mesogen similar to a polycatenar. These observations effectively remove any real demarcation between calamitic and discotic systems.

4.3 Amphitropic Systems

These are materials that form both thermotropic and lyotropic LC phases. Recently Tschierske and colleagues at the Martin Luther University in Halle, Germany have shown that for certain *N*-benzoyl-1-amino-1-deoxyglucitols, related materials, and their mixtures, all the sequential parts of the classical phase diagram of R-theory for lyotropic amphiphile/water systems (I_I, H_I, V_I, L, V_{II}, H_{II}, I_{II}) can be realised thermotropically, dispelling another artificial barrier between lyotropics and thermotropics.

4.4 Other Areas of Fundamental Importance

The length of this overview is limited; the scope of LCs—the fourth state of matter according to some—is immense. Justice cannot be done to all the areas evolving today, and I conclude with a selection of just some of these, well aware that many others are being neglected.

(a) Physical properties, theory, and modeling

Credit must be accorded to the physical scientists who bring so many new techniques (Chandrasekhar 1992; Demus *et al.* 1999), including atomic force microscopy and freely suspended film technology, to bear on LC systems, extending knowledge of their properties and parameters. Their research is fundamental to all in the field, including the theoreticians who, following on those mentioned earlier, continue to effect progress in theoretical understanding of LC systems, importantly by incorporating flexibility, as in real molecules, into the molecules of their models. Important contributions have been made by Chandrasekhar, Madhusudana, and Shashidhar on near neighbor intermolecular correlations and pretransitional effects. Luckhurst and many others have also made big advances in the area of computer simulation of LC systems aimed at the prediction of mesogen behavior through the molecular dynamics approach.

(b) Confined geometry systems

Perfection of thin film LC devices has led to much research on anchoring of mesogens at various surfaces (including gratings), leading to a new branch of the subject dealing with LCs in confined geometries (Crawford and Zumer 1996), also covering LCs in polymer networks, reflective coatings, and anisotropic gels.

(c) Lyotropic systems and biology

Lyotropic LCs, important in oil, food, and detergent technology, also progress strongly with studies of new cationic, anionic, and zwitterionic amphiphiles and increasing interest in carbohydrate-based materials, reminding us of the important role of lyotropics in biological systems and life itself. A valuable review is that by Hiltrop (1994). In this context, other active areas concern the role of hydrogen bonding in creating LC phases, studies of LC supramolecular assemblies, and the synthesis of dendrimers, whereby large spherical assemblies with surfaces which are LC active can be constructed.

(d) Bent or banana-shaped mesogens

Chirality has always been an important feature of LC research and until recently was assumed to be essential for the production of helical LC structures. These can, however, be formed by achiral banana-shaped molecules, and this discovery has led to a wealth of publications in this area.

5. Conclusion

Our understanding of LCs continues to extend and advance through fundamental research stimulated and supported through their technological applications. The cycle is then completed by the feedback of new results from basic research into developing technology for new areas such as in-plane switching displays, cholesteric gel displays, light emitting LCPs, plasma switching of LC displays, and in the area of photonics, spatial light modulators, and optically switched systems. A knowledge of LCs is important today because of their omnipresence, and as already said, the aim of this overview has been to encourage the reader to seek that knowledge from this dictionary and detailed sources such Demus *et al.* (1998).

Bibliography

Brown G H, Shaw W G 1957 The mesomorphic state. Liquid crystals. *Chem. Rev.* **57**, 1049

Carlsson T, Leslie F M 1999 The development of theory for flow and dynamic effects for nematic liquid crystals. *Liq. Cryst.* **26**, 1267–80

Chandrasekhar S 1992 *Liquid Crystals–Second Edition*. Cambridge University Press, Cambridge

Chandrasekhar S, Sadashiva B K, Suresh K A 1977 Discotic liquid crystals. *Pramana* **9**, 471

Crawford G P, Zumer S 1996 *Liquid Crystals in Complex Geometries*. Taylor and Francis, London

De Gennes P G, Prost J 1983 *The Physics of Liquid Crystals*, 2nd edn. Oxford University Press, Oxford

Demus D, Richter R 1978 *Textures of Liquid Crystals*, 2nd edn. VEB Deutscher Verlag fur Grundstoffindustrie, Leipzig, Germany

Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) 1998 *Handbook of Liquid Crystals, Volumes 1, 2A, 2B and 3*. Wiley-VCH, Weinheim

Demus D, Goodby J, Gray G W, Spiess H -W, Vill V (eds.) 1999 *Physical Properties of Liquid Crystals*. Wiley-VCH, Weinheim

Frank F C 1958 Theory of curvature elasticity. *Discuss. Faraday Soc.* **25**, 19–28

Friedel G 1922 Les états mesomorphes de la matière. *Ann. Physique* **18**, 273

Goodby J W, Waugh M A, Stein S M, Chin E, Pindak R, Patel J S 1989 A new molecular ordering in helical liquid crystals. *J. Am. Chem. Soc.* **111**, 8119–25

Gray G W 1962 *Molecular Structure and the Properties of Liquid Crystals*. Academic Press, London

Gray G W, Goodby J W 1984 *Smectic Liquid Crystals*. Leonard Hill, Glasgow, UK

Gray G W, Harrison K J, Nash J A 1973 New family of liquid crystals for displays. *Electron. Letts.* **9**, 130–1

Hilsum C 1991 In: Howells E R (ed.) *Technology of Chemicals and Materials for Electronics*. Ellis Horwood, Chichester, UK

Hiltrop K 1994 In: Stegemeyer H (guest ed.) *Topics in Physical Chemistry: Liquid Crystals*. Steinkopff, Darmstadt and Springer, New York

Kelker H, Knoll P M 1989 Some pictures of the history of liquid crystals. *Liq. Cryst.* **5**, 19–42

McArdle C B (ed.) 1989 *Side Chain Liquid Crystal Polymers*. Blackie, Glasgow, UK

Miyachi K, Fukuda A 1998 In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals*. Wiley-VCH, Weinheim, Vol. 2B

Reinitzer F 1888 Contributions to the knowledge of cholesterol. *Monatshefte* **9**, 421

Schadt M 1994 In: Stegemeyer H (guest ed.) *Topics in Physical Chemistry: Liquid Crystals*. Steinkopff, Darmsatdt and Springer, New York

Schadt M, Helfrich W 1971 Voltage dependent optical activity of a twisted nematic liquid crystal. *Appl. Phys. Lett.* **18**, 127

Serrano J L (ed.) 1996 *Metallomesogens*. Wiley-VCH, Weinheim

G. W. Gray
Wimborne, UK

Liquid Crystalline Polymers: An Introduction

The science of liquid crystals is much older than that of polymers, while the science and practice of liquid crystalline polymers (LCPs) are much more recent than either. To appreciate this comparatively new field, a measure of familiarity with both liquid crystals and polymers is helpful.

The liquid crystalline state is intermediate in nature between a crystal and a normal liquid. A liquid crystal has long-range orientational order between molecules without the three-dimensional long-range positional order which would relate each molecule to a crystal lattice. A liquid crystal thus possesses anisotropy, which is typical of a crystal—optical anisotropy is particularly significant—yet flows like a liquid. If the term “liquid crystal” is too oxymoronic for comfort, then the alternative of “meso phase” is used; this highlights the intermediate character of the substance. On a scale of temperature, the liquid crystal phase is also intermediate in that, on heating, it is revealed at the melting point (T_m) of the crystal phase and occupies a range of stability until the so called “upper” or “clearing” transition. At this temperature, the upper transition temperature or $T_{\text{lc-i}}$, it transforms to the familiar liquid or “melt” phase, which is isotropic. Systems in which liquid crystallinity are seen in a particular temperature range are known as thermotropic.It should also be emphasized that long-range orientational order does not imply that all molecules are perfectly aligned, rather they show preferential orientation about a symmetry axis known as the director. Liquid crystallinity is generally observed in materials formed from molecules which have rod-like or disk-like character. The characteristics of a particular liquid crystal, such as T_m, $T_{\text{lc-i}}$, its optical properties, its thermal stability, etc., can in considerable measure be tailored by appropriate molecular design.

There are three classes of liquid crystal phase: nematic, smectic, and cholesteric. The nematic phase is the most straightforward of the three, in that it contains only long-range orientation order without any long range positional order. Smectic phases contain, in addition, some measure of long-range positional order, which nevertheless falls short of full three-dimensional periodicity. There are numerous classifications of smectic phases, depending on the exact nature of the positional order. Cholesteric phases occur where the forming molecules have a chiral center (carbon atoms where none of the attached groups are equivalent). They are similar in many respects to nematics, yet the molecular chirality imparts a twist about an axis normal to the molecular director. The twist is very gradual at the molecular level, yet it makes a significant imprint on the micron scale and thus on the optical texture. The most important application of liquid crystals is in displays, where their optical anisotropy is allied to their ability to be steered by applied fields, both electrical and surface, to produce the familiar flat panel display devices.

The polymeric state is characterized by long, intertwined, chain-like molecules which are flexible at temperatures above the glass transition. The molecular motion occurs on different scales. There are local motions about flexible links in the chains, which may require a level of local co-operativity. There are co-operative rearrangements of the local packing; these enable motion of segments of the chains and endow polymers with rubbery properties. The viscous flow behavior, typical of conventional liquids, also occurs in polymers. Yet where the molecular weight of the chains is high, viscous flow is very slow and can be eliminated altogether by chemically cross-linking the chains. In a wide range of polymers, crystallization is possible. It usually occurs without eliminating chain entanglements and, because of this compromise, it is normally incomplete. The “crystallinity” of a polymer is an important physical characteristic that has a significant influence on its properties. Except in the crystalline phase, neighboring polymer chains do not lie parallel to each other—although there may be local orientational correlation between some chemical groups (see *Structure of Polymer Glasses: Short-range Order*). Preferred orientation of the chains accompanies mechanical deformation of a polymer, the most important example being the axial alignment of chains on stretching. Such alignment can be stabilized below T_g or by crystallization, but on reheating to the melt or rubbery state, the polymer relaxes, both structurally and mechanically. Polymers, as a result of their long, entangled molecules, are endowed with considerable plasticity and toughness. They consequently have extraordinary utility as structural materials.

1. Main-chain Liquid Crystalline Polymers

1.1 Effect of Molecular Length

A main-chain LCP consists of anisotropic, “mesogenic” units joined end-to-end. Where the individual chemical units are not sufficiently anisotropic in shape to form nonpolymeric liquid crystalline phases by themselves, two questions need to be addressed. First, how long would a rigid rod unit have to be before it became liquid crystalline? Second, if the units are joined together into a chain, how stiff will that chain have to be before it shows liquid crystalline behavior? The first question is really one that is relevant to small molecule thermotropic liquid crystals. Flory (1956) showed, on entropic principles alone, that a rigid rod will switch from the isotropic to the liquid crystalline phase when its axial ratio (length divided by diameter) exceeds 4.0; later work from the same department modified this critical value to 6.4. The introduction of an energy term, which favors parallel packing, may reduce this value. However, it also means that it is possible to simulate an upper transition temperature, which was not possible with the original Flory theory as it was “athermal”. However, the Flory theory also included the effect of dilution with a solvent; this aspect will be discussed in Sect. 1.3 in the context of lyotropic systems.

A question which is more relevant to the stability of a liquid crystalline phase in a polymer system, is how stiff would a chain of anisotropic units have to be before it becomes liquid crystalline? By “how stiff,” the question also implies an additional meaning in this case of “how straight”. A useful measure of straightness is the parameter known as persistence length. This can be understood as the mean distance over which a chain maintains a memory of its starting orientation. For many thermotropic main-chain systems, it has been found that the criterion for liquid crystallinity is that the ratio of persistence length to diameter should exceed 5.0. The fact that the persistence length of a molecule decreases with increasing temperature (as the rotation about chain backbone bonds becomes increasingly easy) means that the critical value of "persistence ratio," as it has been called, also leads to a prediction of $T_{\text{lc-i}}$. The issue of the molecular criteria for liquid crystalline stability in polymers is addressed in more detail by Donald and Windle (1992).

1.2 Managing the Melting Point

The first attempt to make a LCP was by Vorlander in 1923. He synthesized polybenzamide, but found that he had made an infusible solid which, while perhaps birefringent, was not in any sense a liquid. Figure 1 gives an indication of why Vorlander was not successful. It shows data for both $T_{\text{lc-i}}$ and T_m for a series of oligomers made from phenyl groups para-linked through esters. Figure 1 shows that, as expected, the upper transition temperature of liquid crystalline stability increases with increasing length. It also shows that the crystal melting point increases with length. The fact that the melting point increases more slowly than the upper transition means that the temperature range of liquid crystalline stability also increases with length—a fact that would appear to bode well for main-chain polymers. However, at a length of only 8 units, the melting point is approaching 575 K, while for a polymer of perhaps 100 units or more the melting point exceeds 775 K. In a darkened room, the polymer would glow a dull red at this temperature. It leads to chemical breakdown of the polymer as well as being impractical from the standpoint of polymer processing. Many modifications can be made to a chain to reduce the crystal melting point; many of these will reduce $T_{\text{lc-i}}$ as well. Example modifications include the addition of bulky side-groups, short side-chains, or the insertion of increasingly flexible links between the anisotropic units. In addition, the chain can be kinked through the addition of bent units. The price paid for this approach is a reduced temperature range of liquid crystalline stability—a range in which thermal processing must be accomplished if the advantages of a LCP are to be realized.

Another approach is to separate units that are sufficiently anisotropic in shape to be mesogenic in their own right using flexible spacers of increasing length. These flexible spacers are typically sequences of CH_2 units. Table 1 compares a low molecular weight liquid crystalline molecule with the equivalent molecule joined through $-(CH_2)_{10}-$ sequences. (The individual molecule is terminated with two $-(CH_2)_5$ groups). The transition temperatures are of the order of 100 K higher in the connected (polymeric) case. The connectivity reduces the freedom of motion of the molecule in the higher temperature, less-ordered phase, and thus the entropy change at the two transitions. It is also interesting that both the transition temperatures alternate up and down with the incorporation of each additional CH_2 unit into the spacer. This phenomenon, which is known as the "odd–even effect," is attributed to the fact that an odd number of units, as opposed to an even number, has a lowest energy conformation which is kinked rather than straight.

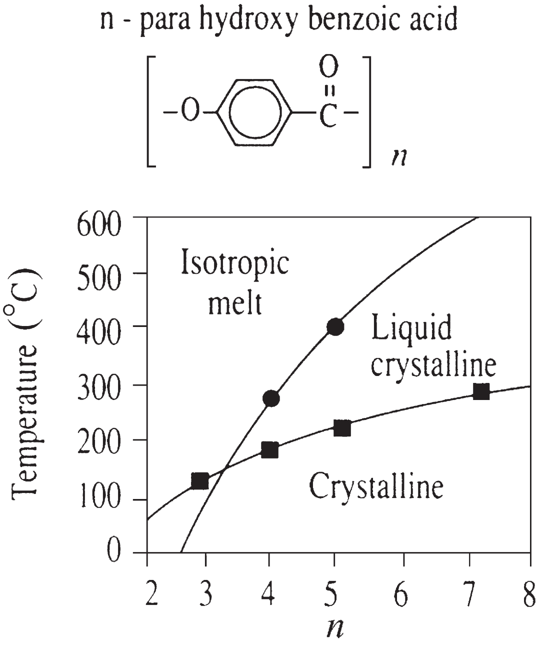
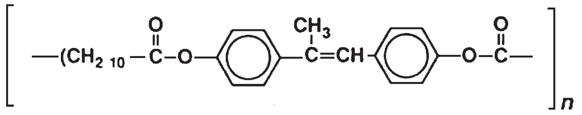
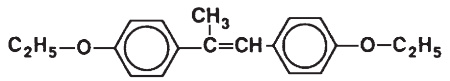


Figure 2
Melting points and upper transition temperatures ($T_{\text{lc-i}}$) for a series of oligomers based on *p*-hydroxybenzoic acid (HBA).

The surest way of decreasing T_m without decreasing $T_{\text{lc-i}}$—and thus avoiding any reduction in the temperature range of liquid crystalline stability—is to build the chain from two or more different units positioned at random to form a random copolymer. The units are typically para-linked. The straightness of the chain is therefore not compromised, although the random positioning of the different units means that the chain cannot contribute to the full three-dimensional periodicity of a crystal lattice. In many of the more widely used thermotropic random copolymer systems, crystallinity is in fact maintained right across the copolymer composition range, even though the crystal melting point is very substantially reduced, especially in mid-range. Figure 2 shows melting point data for the random copolyester system, hydroxy benzoic acid (HBA)–hydroxy naphthoic acid (HNA), the base of many polymers of the Vectra range. Random copolymerisation is by far

Table 1
The influence of connection through flexible spacers to form a chain on the transition temperatures of a small rod-like molecule.

	T_m	$T_{lc\rightarrow i}$
	199 °C	223 °C
	106 °C	121 °C

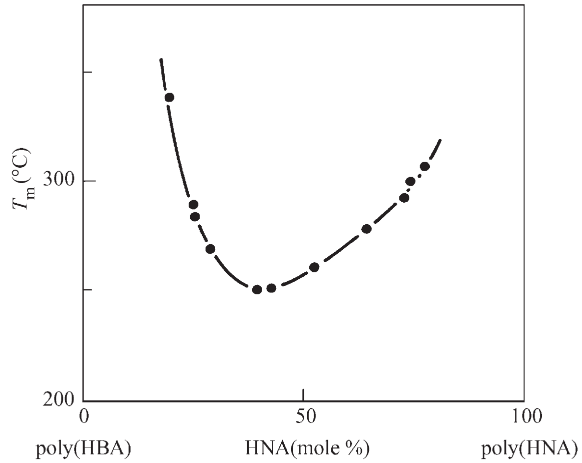


Figure 2
The variation in crystal melting point with composition for a thermotropic random copolyester based on HBA and HNA.

the most widely used mechanism of melting point control in thermotropic main-chain systems.

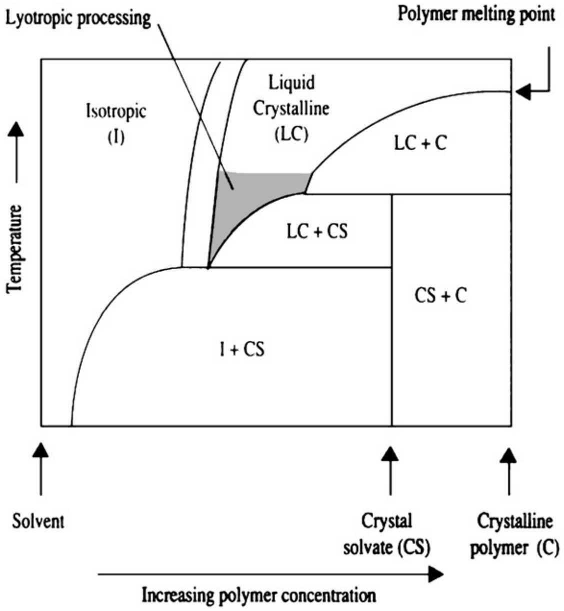

Lyotropic processing
Polymer melting point
Isotropic
(I)
Liquid
Crystalline
(LC)
LC + C
LC + CS
CS + C
I + CS
Temperature
Solvent
Crystal
solvate (CS)
Crystalline
polymer (C)
Increasing polymer concentration

Figure 3
A schematic phase diagram of a rigid chain polymer/solvent system.

1.3 Lyotropic Systems

Systems in which the liquid crystalline phase is x accessed by the addition of a solvent are known as lyotropic rather than thermotropic. The distinction can, however, be overdrawn as solvent additions simply tend to reduce the temperatures bounding the phase in an otherwise thermotropic system. Figure 3 is a schematic phase diagram of a polymer–solvent system containing a liquid crystalline phase. The polymer molecule is intrinsically liquid crystalline in nature (see polymer axis). The addition of the solvent (moving right to left across Fig. 3) reduces the crystal melting point; in practical systems, this means that the system can be processed at more convenient temperatures. As with aspects of molecular design that reduce the melting point, side groups, flexible spacers, etc., $T_{lc\text{-}i}$, is reduced more quickly than the melting point. The temperature range of liquid crystalline stability is therefore reduced and eventually eliminated with increasing solvent content. In (thermodynamic) phase space, where temperature and concentration are both variables, the single phase regions are always separated by a region in which two phases of different composition are in equilibrium with each other. The narrow region separating the isotropic and liquid crystalline phases is known as the "Flory chimney," and was predicted ahead of experimental evidence and even of the recognition of the concept of a LCP. Figure 3 also includes a crystal solvate (CS) phase, which corresponds to a crystal structure in which a proportion of solvent molecules is incorporated into the crystal lattice. It is a feature of some commercially significant systems, particularly those which form the basis of the production of aramid fibers such as Kevlar and Twaron. In these cases, the polymer is poly *p*-phenylene terephthalamide and the solvent is pure sulphuric acid (H_2SO_4). To regain the solid polymer, lyotropic routes require the solvent to be washed out. This limits their applicability to fibers and thinfilms.

1.4 Fridelian Classification: Nematic, Smectic, or Cholesteric

If one envisages main-chain LCPs to be assemblies of smooth snake-like chains, then it follows that the order will be the simplest, namely nematic. With long chains, ends are sparse and it is not sensible to consider that they could segregate into layers in any way equivalent to small-molecule smectics. In any case, there will be polydispersivity—so the chain molecules would be of unequal lengths. However, polymer chains are generally far from smooth and consist of chemical repeats; furthermore there is now evidence (Hanna *et al.* 1993) in some of the more widely used thermotropic copolyester systems, that the parallel chains above the crystal melting point maintain a degree of longitudinal register so that phenyl groups remain opposite phenyl groups on the neighboring chains, and likewise ester linkages with ester linkages. Although the segregation does not involve the chain ends, it is nevertheless a form of smectic-type order. To note the distinction, it has been referred to as "chain smectic." This finding was also significant in that it explained the rapid rate of crystallization observed in these systems; the register is achieved in the highly mobile liquid crystalline phase, leaving only more local rearrangements to be made on subsequent crystallization. Liquid crystalline phases based on alternating mesogenic and flexible spacer sequences also tend to show chained smectic layer structures because the chemically distinct sections of the chain segregate so that like is next to like on neighboring chains. There is evidence that some smectic polymers transform to nematics with increasing temperature, as the one-dimensional positional order is lost before the orientational order is lost. Polymer chains containing chiral centers form the cholesteric phase. The period of the gradual twist is a major contributor to the rich optical texture shown by these polymers. Under conditions where it is, in effect, a diffraction grating for light, the polymers can be most attractive optically.

In general, optical textures are dominated by the presence of topological defects in the liquid crystalline phase. They are essentially singularities in director fields, similar in some respects to the features of a fingerprint. Their exact form depends on the relative magnitudes of the different elastic distortion moduli of the field, which describe the energy penalty of splay, twist, and bend distortions. In main-chain LCPs, the splay constant is particularly high, whereas in smectic structures where the splay of the layers is difficult, the bend constant is higher.

1.5 Properties of Main-chain Liquid Crystalline Polymers

(a) Flow and orientability

Main-chain LCPs show remarkable properties under flow. The first characteristic is that, for a given molecular weight, the polymer has a much lower melt viscosity in the liquid crystalline state than in the isotropic. The reduction can be as much as an order of magnitude. Easier flow of the mesophase may simply be related to the reduced entanglement density of the aligned state. However, for semi-flexible chains such as thermotropic random copolyesters, there is evidence for entanglements at molecular weights above 10000 or so, i.e., the melt shows viscoelastic properties. When such a flow-aligned sample is released mechanically, it retracts in a similar way to a conventional polymer. However, the retraction can crumple the aligned mesophase and lead to banded textures.

Another key aspect of the flow of the liquid crystalline polymeric state is the efficiency of molecular chain alignment under extensional flow, or flows which contain an extensional component. Because there is already local parallelism of neighboring chains and, at least, a reduced entanglement density, the orientation occurs much more readily with strain. It is also possible to orient main-chain material in electrical or magnetic fields; however, flow fields are dominant in terms of application to processing.

(b) Mechanical properties

Oriented LCPs do not contain a distinct amorphous phase unlike conventional polymers. The full elastic stiffness in the direction of the highly oriented chain axes can thus be realized, as can the strength. Fibers drawn in the liquid crystalline phase and subsequently solidified—either by cooling in the case of thermotropics or by removal of the solvent in the case of lyotropics—therefore provide remarkable and very useful mechanical properties. Because they are not limited to thin sections to enable solvent removal, thermotropics can be processed to form bulk materials using injection molding or extrusion. Moldings can show a high degree of molecular alignment in the predominant flow direction, especially near outer surfaces. Such alignment can endow the product with enhanced stiffness and good strength, but only along the flow axis. Tensile stress components normal to this direction can lead to easy splitting. For a closely related reason, weld lines in thermotropic polymer moldings are often sources of considerable weakness, so that optimal mold design is of the essence.

(c) Physical and environmental stability

Thermotropic polymers show little shrinkage on solidification and have a low coefficient of thermal expansion. They are thus excellent candidates for high precision moldings, where it is important that the exact mold shape is reproduced as faithfully as possible in the product. In general, LCPs show remarkable resistance to most solvents—indeed the solvents used in lyotropic processing have to be among the more aggressive of liquids, and are correspondingly cumbersome to handle. The resistance stems from the fact that the segmental increase in entropy on dissolution is restricted, as the additional freedom of motion of the chain in solution is limited by its intrinsic stiffness. Again, this environmental stability is useful in high precision applications where even modest uptake of small molecules would cause swelling and distortion. It also commends the polymer for applications in a wide range of organic environments.

(d) Optical properties

It is interesting that the types of groups (typically ring structures) incorporated into a molecular chain to endow it with stiffness are often those with a tendency to be electro-optically active. Thus the ability to achieve ready alignment in the mesophase means that the active groups can be mutually oriented so as to give the anisotropy essential to many useful optical properties. Furthermore, the high degree of alignment means that the polymer is very homogeneous in terms of density and quality of alignment, and thus scatters little light.

(e) Cost

A significant drawback of LCPs for bulk applications is their comparatively high cost—perhaps 10 times that of a common commodity polyolefin. This mainly stems from the high cost of many of the aromatic monomers used. Molecular engineering is expensive.

2. Side-chain Liquid Crystalline Polymers

2.1 Introduction

While main-chain LCPs were generally developed by polymer scientists who were interested in making their chains stiffer, side-chain materials were developed by people from the small molecule liquid crystalline community who were interested in attaching their molecules to conventional polymer backbones. An early finding was that attachment tended to destroy the liquid crystalline phase. The randomness of the backbone obviously had a dominant influence and compromised the liquid crystalline order which the side groups would otherwise achieve if left to themselves. However, with a flexible spacer between the mesogenic side group and the backbone, the liquid crystalline phase is once again stabilized as the side groups mutually align. This effect is illustrated in Fig. 4. In addition, the temperature range of stability of the mesophase is larger than for the side groups by themselves. Consider for example, the molecule shown in Fig. 5 and a flexible spacer of six CH_2 units. For the mesogenic units by themselves but with the flexible spacer attached as a tail, the liquid crystalline phase is stable over a temperature range of some 20 K, from the crystal melting point of 330 K to $T_{\text{lc-i}}$ at 350 K. With the mesogenic group attached to the acrylate backbone through the spacer of six (CH_2) units, the temperature range of stability increases to 85 K, which is now between the T_g at 305 K and $T_{\text{lc-i}}$ at 390 K. One contributor to this increased range is the suppression of crystallinity as a result of attachment to the backbone. While the backbone eliminated liquid crystallinity when no spacer was present, it has the effect of significantly increasing the upper temperature limit of stability with the spacer in place. The reason for this is not related to the organization of the side chains—which is arguably more difficult as a result of attachment to the backbone—but is due to the restriction of the motion of the side chains in the isotropic phase. The entropy of the isotropic phase is thus reduced and the change in entropy on the loss of liquid crystalline order

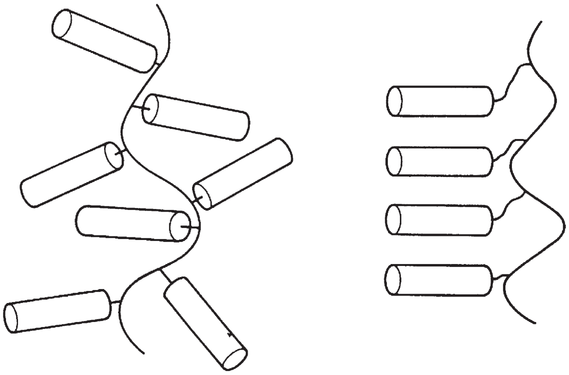


Figure 4
Illustration showing the need for flexible spacers between mesogenic side chains and the backbone to ensure the expression of liquid crystalline order by the side groups.

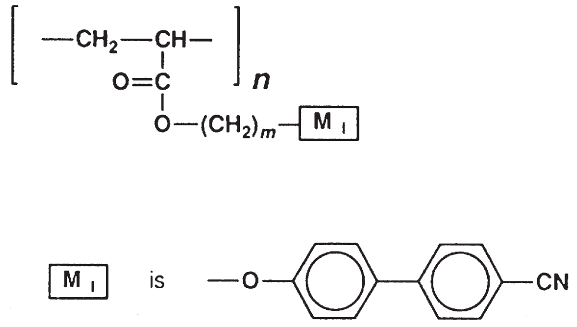


Figure 5
An acrylate backbone with a mesogenic side group and $m(CH_2)$ units forming the flexible spacer.

correspondingly reduced. The result, remembering that $T_c = \Delta H/\Delta S$, is that $T_{lc\text{-}i}$ is increased.

2.2 Friedelian Classification

Side-chain LCP systems lend themselves naturally to smectic structures where the mesogenic groups organize themselves into layers, and the flexible backbone and flexible spacer (and tail if present) are not ordered and fill the space between the layers. In a typical system, the smectic phase transforms to the nematic on heating, and this then transforms to the isotropic. As the amount of flexible material present increases in proportion to the mesogenic groups, the smectic–nematic transition temperature increases, while the nematic to isotropic temperature is reduced. When the temperatures merge with increasing flexible content, the nematic phase is lost; the upper limit of smectic stability begins to decrease as the system becomes increasing dilute in mesogenic groups.

2.3 Control of Structure and Properties in Side-chain Liquid Crystalline Polymers

It is useful to consider the advantages of making a side-chain LCP compared with using just the mesogenic side groups as a low molecular weight liquid crystal. First, the temperature range over which the mesophase(s) is stable is increased. The suppression of crystallinity as a result of attachment to a backbone also has the advantage that optical clarity can be maintained with the polymer thoroughly in the solid state. The second advantage of attaching active groups to a backbone is not unique to the liquid crystalline state. If one wishes to achieve multifunctionality using different types of groups, it is very likely that, as separate molecules, they would segregate. This may be disadvantageous with respect to achieving the right range of properties; certainly in the case of optical properties, segregation is likely to be disastrous as it will render the material opaque. Attachment to backbones in a random fashion encourages functional side groups to remain "in solution." Examples of such an approach would be addition of dye groups, groups with highly nonlinear optical properties, or groups which are cholesteric to those most effective in providing liquid crystalline stability.

Probably the greatest advantage in attaching mesogenic side groups to conventional backbones is the mechanical stability which the backbones endow. If they have a high molecular weight (a condition which is often a special challenge to the synthetic chemist with this class of material), then the viscosity is much increased and stable films may then be produced from them. The active groups may be oriented within these films using external fields, and in some cases, poled to align dipoles to achieve particular properties such as second-order optical nonlinearity. Where the backbone also prevents crystallization and T_g is conveniently placed, the polymer may be cooled to a stable glass film after the orientation process to give an active device that is mechanically stable.

2.4 Liquid Crystalline Elastomers

In this class of side-chain LCP, the conventional backbones are chemically cross-linked to make an elastomer at all temperatures above T_g. A material that is both a liquid crystal and an elastomer opens up a unique series of possibilities for the coupling of stress strain with the orientation of the liquid crystalline texture. The phenomena displayed by such systems, e.g., strain-induced optical switching, intrigue scientifically and challenge commercially.

Bibliography

Donald A M, Windle A H 1992 *Liquid Crystalline Polymers.* Cambridge University Press, Cambridge

Flory P J 1956 *Proc. Roy. Soc.* **234A**, 73

Hanna A, Romo-Uribe A, Windle A H 1993 *Nature* **366**, 546

A. H. Windle
University of Cambridge, UK

This Page Intentionally Left Blank

M

Magnesium: Alloying

Like most metallic elements, magnesium in its pure form is soft and of limited use as a structural material. Besides mechanical properties, many other characteristics can be influenced by adding controlled amounts of other elements to magnesium (ASM 1999). Improvement of processability can be as important as the influence on properties of the final part. In conventional alloy production, alloying is performed by adding elements to molten magnesium, either as pure elements, hardeners, or compounds that are reduced in contact with molten magnesium. A certain liquid solubility is required to form a homogeneous solution at the alloying temperature. Figure 1 shows elements with extensive, limited, and negligible solubility in molten magnesium (Nayeb-Hashemi and Clark 1988, Roberts 1958, Emley 1966). No data have been found for the elements with no indications in Fig. 1. The solubilities are based on binary systems. Ternary- and higher-order effects will in general reduce the solubilities in more complicated alloys. Although most alloying elements are present in substantial amounts, requiring extensive solubility, this is not always the case. Beryllium is active at the ppm level reducing oxidation rate and carbon promotes grain refining even if the liquid solubility is vanishingly small. By applying nonconventional techniques such as stirring-in of particulates, cospraying of different materials, or mechanical alloying/powder sintering, alloys with special compositions can be produced.

The solid solubility of an alloying element is a key factor in determining its effect on properties. The common hardening processes, age hardening, solid solution hardening, and work hardening, rely on a sufficiently high solid solubility. The solid solubility depends on factors like relative atomic size, valency, and electronegativity as well as similarity of crystal structure. A comprehensive discussion of these effects is given by Roberts (1958). Unlike the situation for aluminum, rapid solidification as obtained, for example, in melt spinning, does not cause significant extension of the solid solubility. The following maximum solid solubilities for alloying elements have been reported (at.%) (Nayed-Hashemi and Clark 1988): Ag (3.8), Al (12.0), Ca (0.8), Ce (0.28), Cu (0.15), Gd (4.5), In (19.4), La (0.1), Li (18), Mn (1.0), Nd (0.1), Pb (7.75), Si (0.003), Sn (3.35), Sr($\sim$0), Tb (4.6), Th (0.52), Y (3.75), Zn (2.4), Zr (1.0). As for the liquid solubility, addition of further alloying elements may influence the solid solubility, usually reducing it.

Current magnesium alloys are based on the following alloying elements (in the order of increasing atomic number): Be, C, Al, Si, Mn, Cu, Y, Zn, Zr,

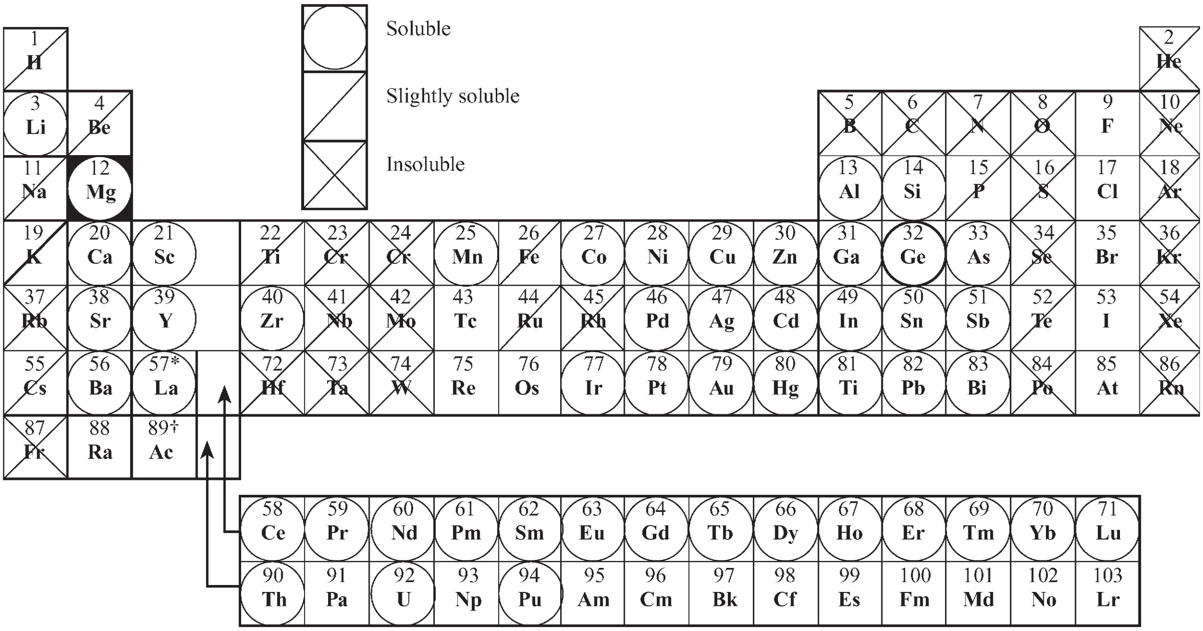

Soluble
Slightly soluble
Insoluble
1 H
2 He
3 Li
4 Be
5 B
6 C
7 N
8 O
9 F
10 Ne
11 Na
12 Mg
13 Al
14 Si
15 P
16 S
17 Cl
18 Ar
19 K
20 Ca
21 Sc
22 Ti
23 Cr
24 Cr
25 Mn
26 Fe
27 Co
28 Ni
29 Cu
30 Zn
31 Ga
32 Ge
33 As
34 Se
35 Br
36 Kr
37 Rb
38 Sr
39 Y
40 Zr
41 Nb
42 Mo
43 Tc
44 Ru
45 Rh
46 Pd
47 Ag
48 Cd
49 In
50 Sn
51 Sb
52 Te
53 I
54 Xe
55 Cs
56 Ba
57* La
72 Hf
73 Ta
74 W
75 Re
76 Os
77 Ir
78 Pt
79 Au
80 Hg
81 Ti
82 Pb
83 Bi
84 Po
85 At
86 Rn
87 Fr
88 Ra
89† Ac
58 Ce
59 Pr
60 Nd
61 Pm
62 Sm
63 Eu
64 Gd
65 Tb
66 Dy
67 Ho
68 Er
69 Tm
70 Yb
71 Lu
90 Th
91 Pa
92 U
93 Np
94 Pu
95 Am
96 Cm
97 Bk
98 Cf
99 Es
100 Fm
101 Md
102 No
103 Lr

Figure 1
Solubility of elements in molten magnesium.

Ag, La, Ce, Pr, and Nd. Earlier alloys containing Li and Th were used for special purposes. Ca, Sr, Pb, Gd, and Tb have been investigated as possible alloying elements.

1. Influence of Alloying

1.1 Molten Metal Reactivity

Magnesium oxide and magnesium nitride are stable compounds, readily formed when molten magnesium is exposed to air. The oxide layer formed on molten magnesium does not form an efficient barrier to further oxidation. Formation of fractures in the surface layer during growth exposes free metal to the atmosphere and accelerated oxidation occurs. This process is also accelerated by the high vapor pressure of magnesium. It has been observed that certain alloying elements reduce the rate of oxidation. Aluminum, which is present in many commercial alloys, significantly reduces the reactivity, possibly because a more dense oxide layer is formed. Beryllium has a profound effect, even at levels of 5–15 ppm commonly used in commercial die casting alloys (Emley 1966). Its strong effect is due to a pronounced enrichment in the surface film.

1.2 Control of Impurity Elements

The growth in the use of magnesium alloys for applications exposed to severe corrosive environments, e.g., automotive parts, relies on the introduction of a series of high-purity alloys for pressure die casting. These alloys contain aluminum as the main alloying element. In preparing such alloys, manganese is first added at relatively high temperatures to a level higher than specified. Following addition of aluminum and sometimes zinc, the temperature of the molten alloy is lowered to the casting temperature. During this period iron will be precipitated together with some manganese and aluminum, as intermetallic constituents. Due to the strong mutual influence of iron and manganese on the liquid solubilities, as well as the decreasing solubility of both elements with decreasing temperature it is possible to reduce the iron content well below 50 ppm (Holta *et al.* 1997). This is required in corrosion-resistant high-purity die casting alloys. Currently no metallurgical ways to remove other critical impurities like copper and nickel have been established. This will be a challenge for the use of magnesium die casting alloys in future closed loop recycling, i.e., recycling back to the original application. Other alloying elements that reduce the iron content are provided by the rare earth elements and zirconium.

1.3 Effect on Microstructure

Magnesium alloys are commonly grouped in two alloy families: (i) magnesium–aluminum alloys and (ii) zirconium grain-refined alloys. The microstructures of these alloy families are easily distinguishable by the grain structure (see Figs. 2 and 3).

(a) Magnesium–aluminum alloys

These alloys develop a dendritic structure with relatively coarse grain size (Fig. 2). Grain refining can be accomplished by stirring-in carbon-containing additives prior to casting. The effect is better the higher the temperature during carbon addition. Formerly hexachloroethane was used; however, due to the health hazards, this has been replaced by wax-fluorspar-carbon mixtures, etc. In pressure die casting, grain refining is accomplished by rapid cooling and turbulent die-filling, hence there is no need for additives. The magnesium-rich part of the magnesium–aluminum phase diagram is shown in Fig. 4. Commercial alloys contain typically 2–10 wt.% aluminum. Zinc in amounts up to 1 wt.% can be added, and in

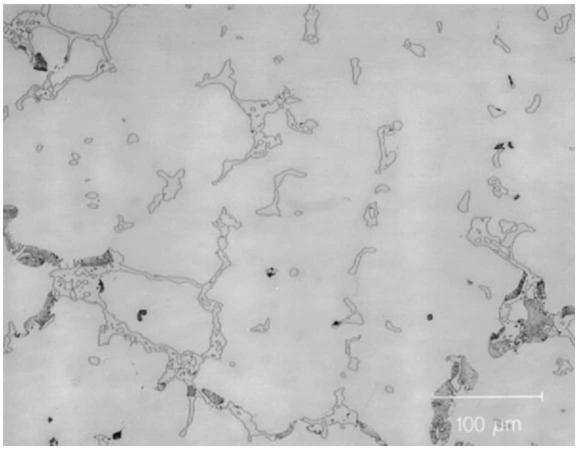


Figure 2
Microstructure of AZ91.

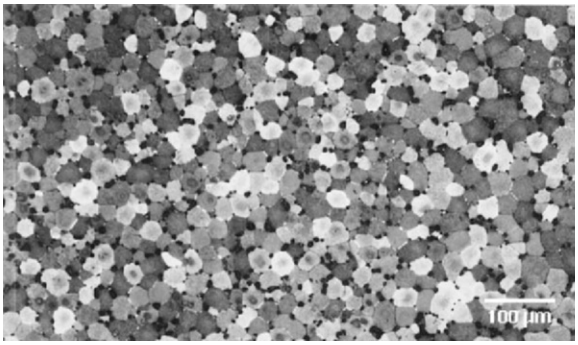


Figure 3
Microstructure of ZE41.

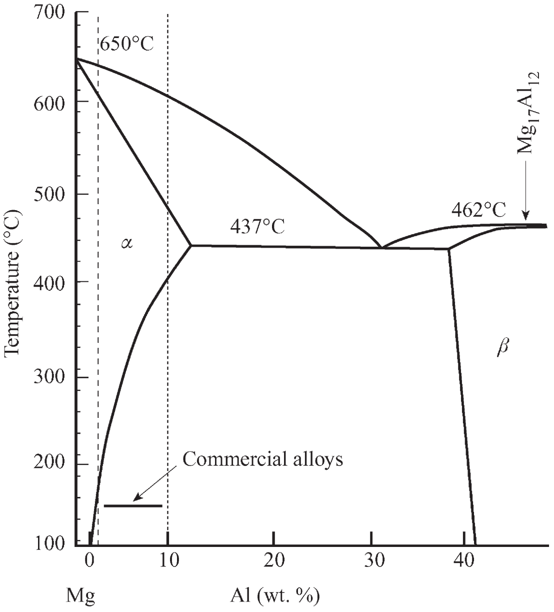


Figure 4
Mg–Al phase diagram.

some instances up to 3 wt.% is used in alloys for sacrificial anodes. Manganese is always added to high-purity alloys to remove iron. Magnesium and aluminum are fully soluble in the liquid state, and a eutectic reaction takes place at 437 °C: liquid→ α(Mg) + β($Mg_{17}Al_{12}$). According to the phase diagram the aluminum contents of the various phases present at the eutectic temperature 437 °C are approximately 33 wt.% (l), 41 wt.% (β), and 12.7 wt.% (α). The lever rule implies that the weight fraction of the brittle β-phase in the eutectic will be approximately 0.72 after solidification. For this reason the amount of aluminum in commercial alloys is limited to a maximum of about 10 wt.%.

During solidification considerable segregation occurs and in practice the Scheil equation, describing solidification under conditions of complete mixing in the liquid phase and no diffusion in the solid phase, provides a reasonable description. The solid solubility of aluminum in magnesium decreases with decreasing temperatures. Solution treatment, T4, is usually carried out slightly below the eutectic temperature, and age hardening is obtained from precipitates formed at temperatures in the range 170–210 °C. The precipitates formed are rather coarse platelets on the (0001)

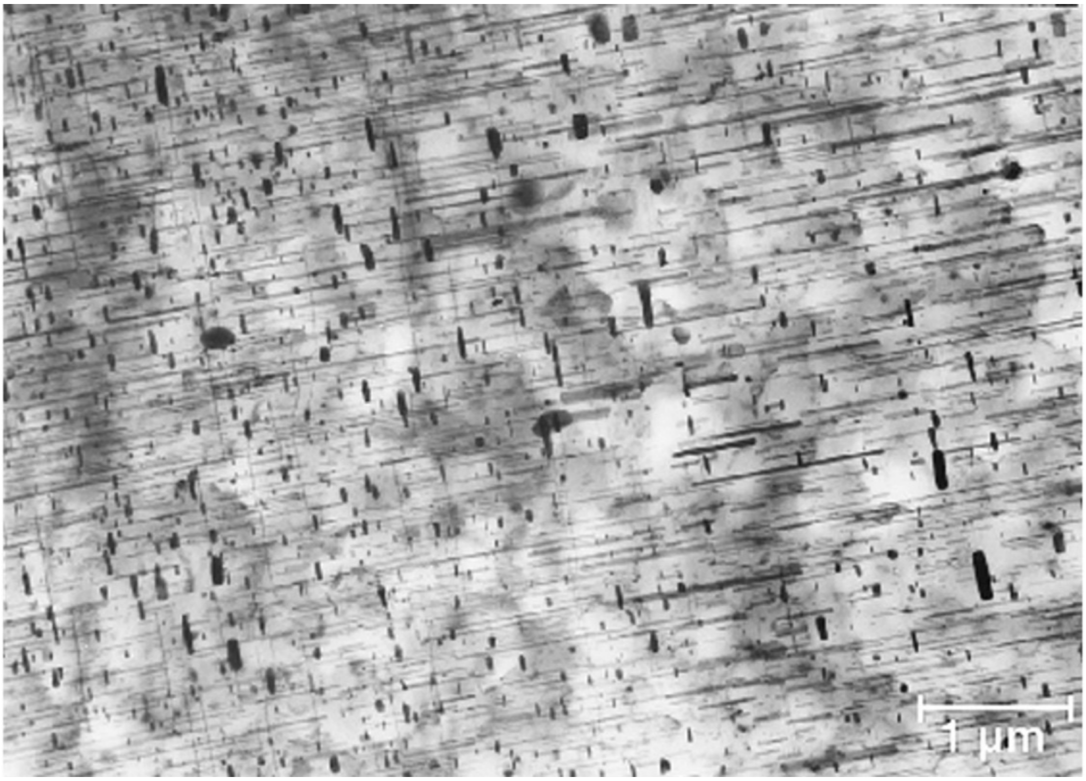


Figure 5
Age hardening precipitates.

planes of the magnesium matrix, and the hardening effect is rather limited (Polmear 1994). Discontinuous precipitation may occur at the grain boundaries, especially at aging temperatures in the range 250–400 °C. For alloys with aluminum content below 6–7 wt.% the hardening effect is negligible.

For castings that are not heat treated, the ductility and fracture toughness increase with decreasing contents of aluminum. This is due to the embrittling effect of the β-phase. For die castings, remarkable ductilities approaching those of wrought materials are obtained at aluminum levels below 6–7 wt.%. Magnesium–aluminum alloys exhibit relatively poor creep properties at temperatures exceeding 120 °C. This is especially the case for fine-grained materials with high levels of aluminum. Excessive creep occurs in the low-melting-point grain boundary regions. Reduction of the aluminum content and addition of silicon, which causes formation of stable Mg_2Si particles, improve creep resistance considerably. Rare earth elements like neodynium, cerium, and lanthanum form intermetallic constituents of the Al_4RE type, adding creep strength both by particle stabilization and by reduction of the volume fraction of the low-melting-point eutectic constituents.

(b) Zirconium grain-refined alloys

Zirconium additions provide an extremely efficient method of grain refining magnesium alloys, Fig. 3. A prerequisite is that elements like aluminum and manganese, which effectively suppress the liquid solubility of zirconium in magnesium, are not present. Elements compatible with zirconium grain refining include zinc, rare earth elements, silver, and yttrium.

While binary magnesium–zinc alloys show inferior mechanical properties and castability, effective grain refinement by zirconium, combined with additions of rare earth elements to reduce microporosity, led to the introduction of alloys like EZ33 and ZE41 (Unsworth 1989). A special alloy, ZE63, provides a high-strength version, where a special solution treatment in hydrogen atmosphere is used to remove some of the embrittling intermetallic phases at the grain boundaries by formation of rare earth hydrides. By this process, zinc is made available for subsequent precipitation hardening during annealing at temperatures around 140 °C. Figure 5 shows a transmission electron micrograph of age hardening precipitates in a magnesium–zinc–rare earth alloy (Wei *et al.* 1995).

Silver promotes efficient age hardening in alloys like EQ21 and QE22. These alloys also contain neodynium-rich mischmetal (a mixture of rare earth elements) and zirconium. The age hardening precipitates are relatively stable at elevated temperatures and provide strength and resistance to creep up to temperatures around 250 °C.

A recently developed commercial alloy system is the WE alloys, containing yttrium and neodynium-rich mischmetal in addition to zirconium. The WE-alloys develop a stable finely dispersed precipitate distribution during heat treatment. The following prerequisites to obtain good high-temperature mechanical strength are fulfilled:

(i) The alloying elements have a high solubility in solid magnesium at elevated temperatures falling sharply with decreasing temperature, allowing precipitation to occur.

(ii) The precipitated particles are stable with a high melting point and corresponding low solid solubilities of the alloying elements, reducing the particle coarsening rate.

(iii) A high amount of magnesium is contained in the strengthening phase, to secure a high volume fraction of precipitates.

The heavy rare earth elements gadolinium and terbium are even more efficient than yttrium in stabilizing the structure, and experimental alloys containing these elements show excellent properties up to 300 °C. As an example a Mg–20 wt.% Tb alloy shows ultimate tensile strength in the range 280–300 MPa and 0.2% proof stress 220–250 MPa at 300 °C (Kamado *et al.* 1992).

Bibliography

American Society of Materials 1999 *Magnesium and Magnesium Alloys*. ASM Specialty Handbook. ASM International, Materials Park, OH

Emley E F 1966 *Principles of Magnesium Technology*. Pergamon, Oxford, UK

Holta O, Westengen H, Røen J 1997 High purity magnesium die casting alloys: impact of metallurgical principles on industrial practice. In: Lorimer G W (ed.) *Proc. 3rd Int. Magnesium Conf.* Institute of Materials, London, pp. 75–88

Kamado S, Iwasawa S, Ohuchi K, Kojima Y, Ninomiya R 1992 Age hardening characteristics and high temperature strength of MgGd and MgTb alloys. *J. Jpn. Inst. Light Met.* **42** (12), 727–33

Nayeb-Hashemi A A, Clark J B 1988 *Phase Diagrams of Binary Magnesium Alloys*. ASM International, Materials Park, OH

Polmear I J 1994 *Light Alloys, Metallurgy of the Light Metals*, 3rd edn. Arnold, London

Roberts C S R 1958 *Magnesium and Its Alloys*. Wiley, New York

Unsworth W 1989 The role of rare earth elements in the development of magnesium base alloys. *Int. J. Mater. Prod. Technol.* **4**, 359–78

Wei L Y, Dunlop G L, Westengen H 1995 Precipitation hardening of Mg-Zn and Mg-Zn-RE alloys. *Metall. Trans.* **26A**, 1705–16

H. Westengen
Norsk-Hydro ISI, Porsvrunn, Norway

Marine Teeth (and Mammal Teeth)

Under the competitive pressures within complex ecosystems, the teeth of marine organisms have developed through evolutionary time to enable the organisms to feed successfully in their particular ecological niche. The diets of organisms vary, in part, due to differences in the susceptibility of their food and to differences in their feeding ability. Such teeth exhibit a diverse range of composition, microarchitecture, and overall morphology as revealed by recent studies of mineralized tissues (Mann *et al.* 1989, Simkiss and Wilbur 1989, Weiner and Lowenstam 1989, Frankel and Blakemore 1991). While many species of marine organisms have teeth constructed predominantly from the polysaccharide chitin (β-1-4-linked polymer of 2-acetamido-2-deoxy-*D*-glucose, often found partially deacetylated), those whose teeth have been modified by the inclusion of inorganic components (biominerals) reveal many of the strategies that have been studied extensively in recent years. Such studies are proving inspirational to materials scientists in their quest for inorganic and composite materials of desired composition, form, and function. An ecological approach provides a particularly helpful perspective to understand the diversity of teeth in marine organisms.

1. *Shark Teeth*

Perhaps the most famous teeth in the marine environment are those of the large predatory sharks, although within the group of sharks and rays, a variety of tooth structures occurs depending on feeding habit. In rays and some sharks, teeth are used mainly for crushing. Light microscope images of teeth from two species of shark are shown in Fig. 1. Their function in this predatory species is to both seize and cut, using an extensive array of such teeth distributed along a readily articulated jaw. New teeth are continually being produced and matured in order to replace those lost during feeding. The hardened outer layer of the teeth, the enamelloid, contains, as its inorganic phase, calcium hydroxyapatite (ideally $Ca_{10}(PO_4)_6(OH)_2$) but with significant substitutions of F^-, Mg^{2+}, and carbonate (CO_3^{2-}) ions, at levels (e.g., as wt.% : F 3.6, Mg 0.34, and CO_3 0.9) generally higher than found in human tooth enamel (Legeros *et al.* 1994). Fish that feed by scraping and excavating coral substrates have specialized teeth with fluoroapatite crystals arranged in bundles throughout the enamel, and often incorporating additional metal ions in the mineralized phases, providing effective toughness for this feeding apparatus. Other species of fish have teeth containing iron on the tooth cap, which are brown in color and are readily detected on the electron microprobe. Possible biological functions for this iron include hardening of the teeth or detoxification by removal of excess iron from the animal (Motta 1987).

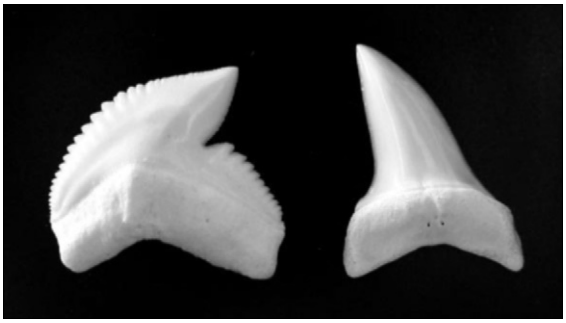


Figure 1
Light microscope image of tooth of (left) the tiger shark *Galeocerda cuvier* and (right) the shortfin mako shark *Isurus oxyrinchus.* The teeth are ~2 cm in their longest direction (specimens courtesy of Museum of Western Australia).

2. *Functional Group Approach*

A particularly insightful analysis of the relationships between tooth structure, use in feeding and ecological niche is provided by the functional group analysis of herbivorous molluscs (Steneck and Watling 1982). Thus, algas are graded in seven groups according to their grazing difficulty or toughness, ranging from microalgas to filamentous algas and on to calcareous algas and crustose coralline algas (the most difficult to graze). The teeth of the molluscs that graze on these algas are arranged on a tongue-like organ (the radula) but differ greatly in their design and microarchitecture while providing excavation ability, contact points and effective width of the graze stroke appropriate for feeding on the desired algas. In one design, many teeth are employed in a sweeping action (“brooms”) but the teeth are not mineralized or hardened, consisting mainly of the polysaccharide chitin. Similarly constituted teeth are found in another group of molluscs whose teeth (“rakes”) are used to graze filamentous algas. A dramatic change occurs in the next group of herbivorous molluscs, the limpets, whose radula has a greatly decreased number of teeth per row but the major teeth (“shovels”) are hardened by the inclusion of silica and iron biominerals. These teeth provide greater excavating ability on harder substrata such as leathery macrophytes and crustose coralline algas. In the final group, the chitons, teeth for excavation (‘multi-purpose tool’) can have up to three cusps, also heavily mineralized using iron and calcium biominerals, with soft unmineralized marginal teeth that lightly sweep the substratum. In this analysis, the feeding apparatus of the chiton is the most versatile of all the groups since it can sweep and excavate simultaneously.

3. Limpets and Chitons

The early identification by Lowenstam (1962, 1967, 1971) of the dominant minerals present in limpet and chiton radular teeth has stimulated extensive studies of these teeth, through which some of the prominent paradigms of biomineralization have been enunciated (Lowenstam 1981). A schematic cross-section of the anterior end of a chiton is shown in Fig. 2. The major features associated with the radula are shown although, for clarity, the radular muscles are omitted. The anatomical structures in the limpet are very similar. The radula consists generally of over 70 rows of teeth ranging from young, soft, unmineralized teeth (still enclosed in the radular sac whose cells deliver ions for mineralization) to mature, hardened, and heavily mineralized teeth that are used in feeding. Thus, in a single radula, there are teeth at all stages of mineralization that can be used in identifying and characterizing the processes of mineralization.

3.1 Limpets

The radula teeth of a typical limpet (*Patella laticostata*) and a chiton (*Acanthopleura hirtosa*) are shown in Fig. 3. In both cases, the cusp and major lateral teeth strike the substratum during feeding and are hardened by the inclusion of iron biominerals. Extensive studies (Webb *et al.* 1989) reveal that, in the limpet mineralized teeth, long acicular needle-like crystals (0.2–1.0 µm in length, 15–20 nm in width) of the iron biomineral, goethite (α-FeOOH) are embedded in a fibrous matrix oriented generally with the [001] needle axis. These radula teeth have a characteristic brown color. The crystals are remarkable in being single domain crystals as shown by high-resolution electron microscopy and by ^{57}Fe Mossbauer spectroscopy, exhibiting very narrow spectral line widths. Another, minor phase of iron present in the teeth (and detectable by Mossbauer spectroscopy) is a poorly crystalline superparamagnetic material (nanoscale particles of iron(III) oxyhydroxide) accounting for about 6% of the Mossbauer spectral area in the mature teeth and located predominantly in the base of the teeth (almost all of the goethite is located in the cusp). The other biomineral present in limpet teeth is silica (SiO_2), which is observed in a range of structural motifs (globular, silicified hollow tubes, and silicified fibrous organic framework) constructed from aggregates of 5–15 nm amorphous particles. The silica particles are amorphous under electron microscopy but distinct chemical environments for Si can be observed by solid state nuclear magnetic resonance spectroscopy. The tooth contains regions of distinct microarchitecture, with the silica region located in the tooth cusp, behind the goethite-containing region. Other minor elements identified in the tooth may be involved in matrix modification, such as the putative role for Cu in enzyme-mediated oxidative cross-linking of the organic matrix.

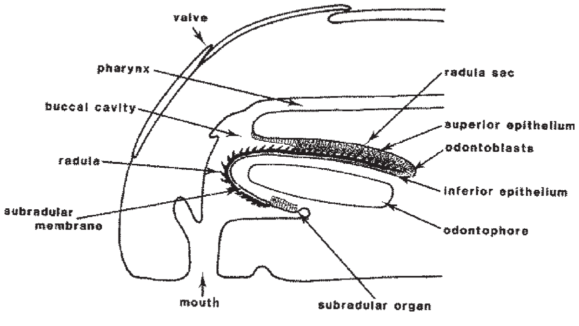


Figure 2
A schematic cross-section of the anterior end of a chiton.

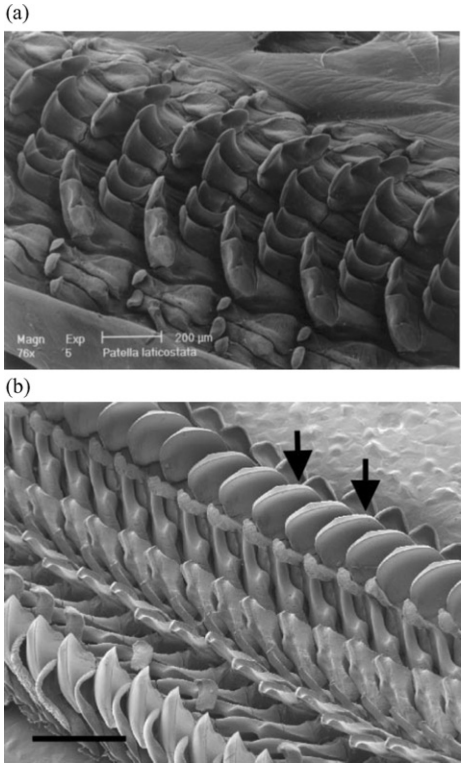


Figure 3
Scanning electron microscope images of the radula of (a) the limpet (*Patella laticostata*) and (b) the chiton (*A. hirtosa*) where arrows indicate the major lateral teeth. Scale bar is 500 µm.

3.2 Chitons

The dominant feature of the major lateral teeth of the chiton radula is their blackening by the presence of the iron biomineral magnetite (Fe_3O_4) on the cusp and extending along the posterior side used for feeding (for review see Webb *et al.* 1989). Depending on the species, other biominerals identified in the teeth include a carbonated hydroxyapatite and fluoroapatite as well as poorly crystalline iron phases, some with phosphate. The elemental composition of the teeth has been studied by diverse analytical techniques such as proton-induced x-ray emission and under the electron microprobe. Mapping of teeth at different stages of mineralization has indicated that elements are mineralized at different stages. Iron appears early on, leading to a rapid blackening of the teeth. Calcium, fluoride, and phosphorus are detected at a later stage (for the species *A. hirtosa*) corresponding to the appearance of a fluoroapatite biomineral. A distinctive increase in Zn appears during mineralization in this species, corresponding perhaps to the hardening of the organic matrix through secondary cross-linking.

The first mineralization product identified in the chiton tooth is electron-dense nanoscale particles whose electron diffraction pattern is that of ferrihydrite ($5Fe_2O_3.9H_2O$). Similar particles are observed in the superior epithelial cells, arranged within aggregates known as *siderosomes* (1 μm and smaller diameter), through which the mineralizing ions are transported. The rate and mechanistic details of this transport process and transmembrane transfer are not known. Extensive and complex mineralization is evident by the following tooth row. This is in marked contrast to that observed in the limpet where a single crystalline phase, goethite, dominates within the tooth cusp. In the chiton, in addition to ferrihydrite, there can be three other iron biominerals within an individual tooth: goethite, lepidocrocite (γ-FeOOH), and magnetite. The magnetite quickly dominates the iron biominerals formed such that, as shown by Mossbauer spectroscopy, it accounts for almost all of the spectral area. The distinctive pair of magnetic sextets corresponding to the two sublattices of magnetite is apparent but the spectrum also contains a doublet, corresponding to 15% of the spectral intensity, which is due to an iron(III) component. Furthermore, in some studies, oxidation of magnetite to maghemite (γ-Fe_2O_3) has been reported.

Ultrastructural studies of the more heavily mineralized teeth are severely hindered by difficulties in sectioning. Access to inner regions of these teeth has been achieved through embedding the teeth in resin and then carefully grinding them down. This procedure provides smooth surfaces for electron microscopy and Raman vibrational spectroscopy, allowing simultaneous mapping of elemental content and mineral phases. For the species *A. hirtosa* the results of this mapping in a mature tooth are shown in Fig. 4. The complexity of the mineralization is apparent, with three iron oxide phases being identified as well as the apatite. The control of mineralization to create these distinct regions of mineralization is thought to involve both differential cellular processes as well as *in situ* factors related to the organic matrix. This matrix contains polysaccharide and protein. For the chiton *A. hirtosa*, the chitin is α-chitin (consisting of arrays of antiparallel hydrogen-bonded chains) and the protein is rich in the acidic side chains aspartic and glutamic acids, with phosphoserine also being detected in appreciable quantities. These are likely nucleation sites for calcium mineralization. The chitin and protein components possess a close spatial relationship, with a number of factors suggesting that the degree of organic-matrix mediated control differs in the different regions of the tooth. Furthermore, the physical arrangement of the matrix fibers should contribute to the mechanical properties of the tooth: the magnetite—mineralized posterior region is hardened appropriately for abrading the rocky substrate and its constituent crustose coralline algas, while the matrix-rich apatite and lepidocrocite-mineralized regions are constructed so as to absorb the physical shocks encountered during the feeding process.

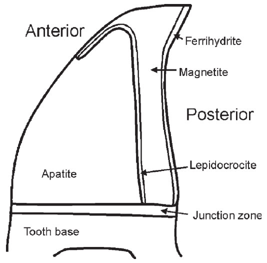


Figure 4
Schematic cross-section of mature tooth of the chiton, *A. hirtosa*, showing the distribution of phases.

4. Concluding Remarks

Biology provides a rich, exotic array of feeding strategies within the marine environment. The present article does not seek to be comprehensive. For example, a particularly interesting modification of the radula for feeding is provided by the cone shells which use a specialized harpoon to immobilize prey (Olivera *et al.* 1985). The harpoon is essentially a hollow radula tooth and the harpoons are stored in the radula sac, analogous to the anatomical features of the chitons and limpet shown in Fig. 1. Finally, the characteristics of the teeth of parasitic lamprey species have been correlated with their feeding habits, particularly the preference for flesh-feeding or blood-feeding (> Potter and Hilliard Potter and Hilliard 1987).

See also: Shell: Properties

Bibliography

Frankel R B, Blakemore R P 1991 *Iron Biominerals*. Plenum Press, New York

Legeros R Z, Bautista C, Wong J L, Legeros A, Valdecanas A, Legeros J P 1994 Biological apatites in modern and fossil shark teeth and in calcified fish scales. *Bull. de l'Institut oceanographique* **14**, 229–36

Lowenstam H A 1962 Goethite in radular teeth of recent marine gastropods. *Science* **137**, 279–80

Lowenstam H A 1967 Lepidocrocite, an apatite mineral and magnetite in teeth of chitons (Polyplacophora). *Science* **156**, 1373–5

Lowenstam H A 1971 Opal precipitation by marine gastropods (Mollusca). *Science* **171**, 487–90

Lowenstam H A 1981 Minerals formed by organisms. *Science* **211**, 1126-1111

Mann S, Webb J, Williams R J P 1989 *Biomineralization: Chemical and Biochemical Perspectives*. VCH, Weinheim, Germany

Motta P J 1987 A quantitative analysis of ferric iron in butterfly fish (*Chaetodontidae perciformes*) and the relationship to feeding ecology. *Can. J. Zool.* **65**, 106–12

Olivera B M, Gray W R, Zeikus R, McIntosh J M, Varga J, Rivier J, de Santos V, Cruz L J 1985 Peptide neurotoxins from fish-hunting cone snails. *Science* **230**, 1338–43

Potter I C, Hilliard R W 1987 A proposal for the functional and phylogenetic significance of differences in the dentition of lampreys (Agnatha: *Petromyzontiformes*). *J. Zool.* **212**, 713–37

Simkiss K, Wilbur K W 1989 *Biomineralization: Cell Biology and Mineral Deposition*. Academic Press, London

Steneck R S, Watling L 1982 Feeding capabilities and limitation of herbivorous molluscs: a functional group approach. *Marine Biol.* **68**, 299–319

Webb J, Macey D J, Mann S 1989 *Biomineralization*. In: Mann S, Webb J, Williams R J P (eds.) *Chemical and Biochemical Perspectives*. VCH, Weinheim, Germany, pp. 345–87

Weiner S, Lowenstam H A 1989 *On Biomineralization*. Oxford University Press, Oxford

L. R. Brooker, A. P. Lee, D. J. Macey, and J. Webb
Murdoch University, WA, Australia

Martensite

The phase termed martensite in ferrous alloys is the product of a displacive, diffusionless transformation of the high-temperature f.c.c. solid solution austenite (γ), cooled at a rate sufficiently rapid to suppress diffusional decomposition at temperatures above the martensite range. In carbon steel, the martensite produced by quenching may thus be regarded as a supersaturated solid solution of interstitial carbon in b.c.c. iron, with a structure that is a distorted (tetragonal) version of the equilibrium ferrite (α) structure (and commonly designated α'). Martensitic reactions are also observed in a number of interstitial-free substitutional ferrous alloys (Fig. 1), the product structure being of cubic symmetry if the parent phase is an ideal solid solution and tetragonal (t) if the martensite inherits some degree of order from the austenite. In certain systems, with an austenite of low stacking fault energy (SFE), a martensitic transformation to a fully coherent, close-packed hexagonal (c.p.h.) product (ε martensite) is observed. In most cases, formation of ε martensite precedes formation of α' martensite, but it may also form within the strain fields associated with α' martensite.

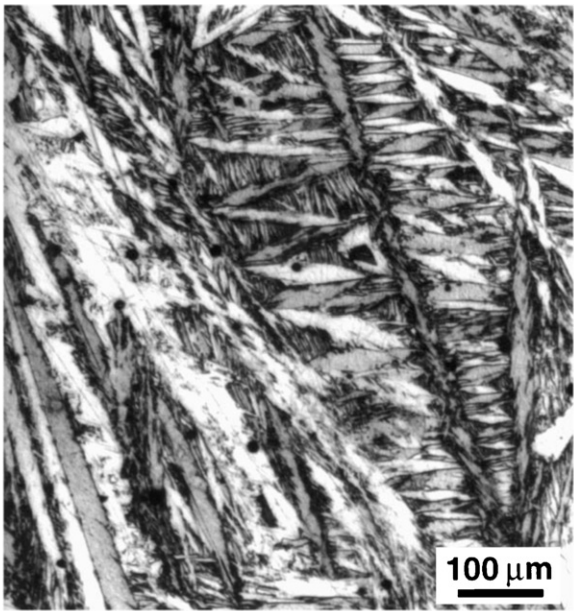


Figure 1
Reflected light micrograph of martensite that is the product of "burst" transformation in an Fe–30 wt.% Ni alloy. The zigzag pattern of plates is considered the product of autocatalytic nucleation, the nucleation of one martensite variant being stimulated in the strain field of a preceding plate.

1. *Thermodynamics*

In a broad range of ferrous alloys, cooled rapidly from temperatures in the austenite phase field, the diffusional decomposition of the austenite may be suppressed. If the austenite that is stable at elevated temperatures is retained on cooling to temperatures significantly below that at which the free energies of parent and product phases are equal (T_0), then there is a chemical driving force for the formation of a metastable martensitic product (α') with a composition equivalent to that of the austenite from which it would form. Martensitic transformation commences on cooling at a temperature (M_s) for which the level of undercooling below T_0 and thus the chemical driving force are sufficient to overcome the energy barrier to martensite nucleation that arises from a combination of interfacial and strain energies. The transformation is reversible on reheating but, for most ferrous alloys, the reverse transformation does not commence until a temperature (A_s) that is typically well above T_0 is reached and there is often a large thermal hysteresis.

The transformation temperatures M_s and A_s are typically strongly dependent on composition in both substitutional and interstitial alloys of iron, with an increase in solute content invariably leading to a decrease in both these critical temperatures. In the absence of rigorous thermodynamic analysis, the equilibrium temperature is commonly approximated as $T_0 = \frac{1}{2}(M_s + A_s)$, while empirical relationships have been developed to provide an approximate guide to the influence that alloying elements have on the transformation temperatures (e.g., Honeycombe and Bhadeshia 1995). The origins of the composition dependence M_s are not fully understood. However, the role of alloying elements in strengthening and thus stabilizing the austenite mechanically has been established as a key factor, and there is evidence that M_s is also influenced by the degree of order in the austenite, the volume and symmetry changes that accompany transformation, and, where appropriate, variations in the magnetic properties of the parent phase.

Given that the transformation is displacive, application of stress during transformation may assist or oppose the transformation depending on the nature of the stress. For a polycrystalline sample subjected to a uniaxial tensile stress, the transformation may be stress-activated at temperatures above M_s. Within the temperature range M_s up to a limiting M_d, martensite nucleation may be either stress-assisted through a positive interaction between the applied stress and the transformation shear, or strain-induced through the introduction of appropriate nucleating defects during plastic deformation of the austenite.

2. *Crystallographic Characteristics*

At the structural level, a martensitic transformation is characterized by the maintenance of a lattice correspondence, which associates uniquely a unit cell of the parent lattice with a unit cell of the product. The Bain correspondence between f.c.c. austenite and b.c.c.(t) martensite (Fig. 2) identifies a b.c.t. cell of axial ratio $\sqrt{2}$ within the f.c.c. structure, with a cube axis becoming the tetragonal c axis and two cubic $\langle 110 \rangle$ directions normal to this defining the remaining $\langle 100 \rangle$ axes of the tetragonal cell. This correspondence implies that the change in structure may be accomplished by atomic displacements equivalent to a homogeneous deformation (or Bain strain) of the parent lattice, and the symmetry of the martensite is determined by the point symmetry of the parent austenite and the symmetry of the Bain strain. Except for certain degenerate cases, the symmetry elements that survive the strain form a subgroup of the symmetry elements of both the parent lattice and the strain. In most instances, the symmetry of the martensite structure is thus lower than that of the parent phase and that of the low-temperature equilibrium phase it most closely resembles. In ferrous martensite, a tetragonal structure is most commonly sustained in those systems in which the martensite inherits point defect arrays or local or long-range order present in the austenite.

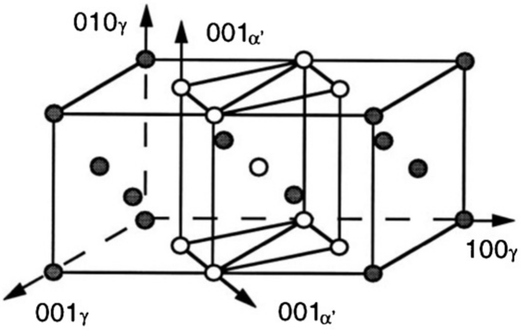


Figure 2
Schematic representation of the Bain lattice correspondence between principal axes of b.c.t. martensite (α') and f.c.c. austenite (γ) unit cells. The Bain strain involves a contraction of $\sim$20% along $[001]_{\alpha'}$ and expansions of $\sim$12% parallel to $[100]_{\alpha'}$ and $[010]_{\alpha'}$ to generate an α' cell of appropriate dimensions.

For the martensitic f.c.c. to b.c.c.(t) transformation, an orientation relationship is invariably observed between the parent and product structures. This relationship varies with alloy composition over a relatively narrow range between two simple extremes, commonly referred to as the Kurdjumov–Sachs (K-S) and Nishiyama–Wasserman (N-W) relationships, and described by

$$\begin{array}{llll} \text{K-S}: & (101)_{\alpha'} \| (111)_{\gamma} & \text{N-W}: & (101)_{\alpha'} \| (111)_{\gamma} \\ & [11\bar{1}]_{\alpha'} \| [1\bar{1}0]_{\gamma} & & [10\bar{1}]_{\alpha'} \| [1\bar{2}1]_{\gamma} \end{array}$$

These two relationships are related by a rotation of approximately 5° about $[101]_{\alpha'}$. Measured orientation relationships are in practice irrational and neither the close-packed planes nor the close-packed directions they contain are exactly parallel.

In ferrous alloys, the martensite phase adopts a plate-like (Fig. 1) or, in some instances, lath morphology, and there is a change in shape which accompanies transformation and which occurs homogeneously over the transformed volume. This shape change has the geometric characteristics of an invariant plane strain (IPS), in which the interface plane (the habit plane) remains invariant (i.e., undistorted and unrotated), thus minimizing the strain energy associated with the homogeneous deformation of a macroscopic region of the parent crystal. Evidence for this change of shape is most readily observed at a free surface where the intersection of a martensite plate with the surface leads to a uniform tilting of the surface and pronounced surface relief. For the f.c.c. to b.c.c.(t) transformation, the magnitude of the shape strain is typically 0.2, while the volume change is of the order of 3% (0.03). The shape strain thus comprises a large component of shear on the habit plane with a small dilation normal to the habit plane.

In the case of the f.c.c. to c.p.h. transformation, the product ε martensite is coherent with the parent austenite and the habit plane is rational $\{111\}_\gamma$. However, for the f.c.c. to b.c.c.(t) transformation, the habit plane, whilst generally regarded as unique for a given alloy, is most commonly irrational and the interface between martensite and austenite is partially coherent. Measured habit planes have been observed to vary with alloy composition and M_s temperature, and with the strength and stacking fault energy of the parent austenite. For those plate martensites forming at low temperatures (0.6–1.4 wt.% carbon in plain-carbon steel), the measured habit planes are typically described as being close to $\{3,10,15\}_\gamma$ or $\{259\}_\gamma$, while at intermediate transformation temperatures the habit planes are more commonly close to $\{225\}_\gamma$. At elevated transformation temperatures (<0.6 wt.% carbon in plain-carbon steel), the product is lath-like and the habit plane approaches $\{111\}_\gamma$.

3. *Crystallographic Theory*

If the homogeneous lattice strain **B** implied by the lattice correspondence is combined with an appropriate rigid body rotation **R**, the matrix product defines a homogeneous total lattice strain $\mathbf{S_t}$ $(=\mathbf{RB})$ that will generate the martensite lattice in its observed orientation relationship with the parent lattice. However, this total strain rarely leaves a matching plane between parent and product structures that is unrotated, and it is thus incompatible with a macroscopic shape deformation that is an invariant plane strain.

This apparent anomaly may be reconciled if it is assumed that the total strain is only locally homogeneous and occurs inhomogeneously on a macroscopic scale so as to maintain an interface plane that is undistorted. The shape strain $\mathbf{S_r}$ may thus be considered the product of the total strain and an additional strain, which periodically relieves the accumulating misfit across the interface in order that the habit plane remain invariant, i.e., $\mathbf{S_r}=\mathbf{S_t}\mathbf{L}$. Since it can produce no further change in structure, the strain **L** is lattice invariant and, in the simplest form of the theory, is assumed to constitute either a slip or a partial twinning shear in parent or product lattices. Alternatively, the total strain $\mathbf{S_t}$ may be considered the product of consecutive invariant plane strains, i.e., $\mathbf{S_t}=\mathbf{S_r}\mathbf{L}^{-1}$, and is thus an invariant line strain, with the invariant line parallel to the line of intersection of the habit plane and the plane of lattice invariant shear (LIS).

This mathematical description of the transformation crystallography forms the basis of the phenomenological theory of martensitic transformations. Given as input the lattice correspondence and lattice parameters of initial and final lattices, and the plane and direction of LIS, the theory permits calculation of the magnitude, plane, and direction of the shape strain $\mathbf{S_r}$, the magnitude of the LIS, and the orientation relationship between the lattices. If $\mathbf{S_t}$ is itself an invariant plane strain then the analysis is simplified; there is no requirement for an LIS and the two lattices related by **B** may be fully coherent across a planar interface. One such transformation is the f.c.c. to c.p.h. transition in austenite of low SFE, where the interatomic spacings within parallel close-packed planes are similar and the change in structure may be accomplished by a shear, equivalent to one half of the twinning shear, between alternate pairs of $\{111\}_\gamma$ planes in any one of three $\langle 11\bar{2}\rangle_\gamma$ twinning directions.

For transformations for which $\mathbf{S_t}$ does not allow an invariant habit plane, $\mathbf{S_t}$ may be combined with a lattice invariant strain **L** to define an invariant habit plane, across which there is sufficient structural continuity to permit rapid glissile motion of the semicoherent interface. Predictions of the crystallographic theory, based on the assumption that the LIS is a simple shear, are in good agreement with available experimental measurements for many transformations, including those ferrous martensites with a habit plane near to $\{3,10,15\}_\gamma$. However, in this form the theory is incapable of accounting satisfactorily for the variations in crystallography observed in ferrous alloys, most notably for those transformations of $\{hhl\}_\gamma$ habit. These include plate-shaped transformation products with $\{225\}_\gamma$ habit plane and lath martensites, where the habit plane moves towards $\{111\}_\gamma$.

4. *Morphology and Substructure*

Ferrous martensites with habit planes near to $(3,10,15)_\gamma$ take the form of thin, parallel-sided plates that are completely internally twinned on $(112)_{\alpha'}$. As the habit plane deviates from $(3,10,15)_\gamma$, the martensite plates become progressively less regular and the substructure more complex. In Fe–Ni alloys, as the weight fraction of nickel is reduced from 0.35 to 0.29, the plates adopt a more lenticular form (Fig. 1), the twin density decreases, and the twins do not extend completely across a plate. Martensite plates of $(225)_\gamma$ habit typically have one near planar interface with the surrounding austenite and one less regular interface, decorated with small subsidiary sideplates. Pairs of plates are often described to form in a ''butterfly'' or ''angle-profile'' configuration. The substructure of $(225)_\gamma$ plates is complex; while $(112)_{\alpha'}$ twins are observed in most plates, the density and extent of twinning vary significantly from plate to plate and there are other planar inhomogeneities within the plates. As the morphology and substructure of the martensite plates become less well defined, the agreement between predictions of the crystallographic theory, based on a simple LIS, and experimental observations becomes progressively less satisfactory.

The transition to a lath morphology that occurs in low-carbon martensite and in certain substitutional ferrous alloys remains incompletely understood. A lath product is common in transformations occurring at elevated temperatures and in systems in which both parent austenite and product martensite have relatively low resistance to plastic flow. There is also evidence of isothermal formation of lath martensite in, for example, Fe–Ni–Mn alloys at temperatures below ambient. In each case, transformation involves growth of thin, parallel laths in ''packets'' in which each adjacent lath has the same shape deformation and orientation relationship with the parent austenite. Except in instances in which adjacent lath pairs are twin-related, thin films of retained austenite are commonly observed between laths and between lath packets, implying that each lath within a packet is independently nucleated. Individual laths typically have a dense dislocation substructure, which is considered to be the product of extensive plastic accommodation of the shape strain within the martensite.

5. *Martensite Nucleation and the Kinetics of Transformation*

Martensitic transformations are first-order transformations and thus proceed via the localized nucleation and growth of individual crystals of martensite within austenite. Martensite plates typically form on multiple variants of the habit plane in a given grain of austenite. In some systems, nucleation may occur isothermally as the result of thermal fluctuations at temperatures above the range in which athermal nucleation occurs, but in ferrous alloys nucleation is more commonly athermal and proceeds with reduction of temperature below M_s.

Theoretical approaches to martensite nucleation are commonly classified in two groups: those that derive from conventional models of homogeneous or heterogeneous nucleation and those that involve considerations of lattice instability. In the classical approach, the strain energy accompanying nucleation, which has both shear and dilational components, is minimized for a disk-like form, the scale of which is determined by a combination of strain energy and interfacial free energy. Reasonable estimates of the energy barrier to nucleation rule out any possibility of homogeneous nucleation and this has led to various models for heterogeneous nucleation in association with some form of critical nucleating defect. The importance of such defects to nucleation has been confirmed experimentally through direct observation and experimental studies of transformations occurring in small particles and coherent precipitates.

Defects comprising suitably spaced groups of dislocations or faults have been shown capable, in principle, of reducing the energy barrier (and thus the critical nucleus size) to thermally activated (isothermal) nucleation. However, there is little direct experimental information on the nature of the defects that in practice facilitate nucleation. In certain circumstances, the defect configuration is such that a temperature may be reached where the activation barrier to heterogeneous nucleation vanishes, allowing athermal nucleation to proceed. A special example involves nucleation of fully coherent ε martensite, where the lattice deformation is an IPS and the interfacial energy is correspondingly small. Transformation is preceded by dissociation of lattice dislocations into widely spaced partial dislocations separated by a layer of stacking fault with a configuration approximating to a thin layer of the c.p.h. phase. The critical nucleus is thus deemed to comprise either a single layer fault or a number of such faults on alternate $\{111\}_\gamma$ planes.

An alternative, nonclassical approach to martensite nucleation arises from consideration of potential localized mechanical instability in the vicinity of a lattice defect. Some displacive transformations, involving small lattice strains, may be interpreted in terms of a general lattice instability (soft phonon mode) that manifests as a static wave representing a structural change throughout the parent crystal. However, with their relatively large lattice strains, martensitic transformations in ferrous alloys are unlikely to occur in this manner.

Once nucleated, growth of individual martensite plates in most austenites is so rapid that the global transformation kinetics can be considered to be controlled exclusively by the rate of nucleation. Interface velocities approach one half of the velocity of a transverse shear wave in the parent crystal. Individual plates partition the volume available to them, apparently reaching a size determined by mechanical constraints to growth. In most ferrous alloys (including plain-carbon and low-alloy steels), where the transformation strains are large and there is no observable lattice softening through the transformation range, plastic accommodation accompanies formation of even very small plates. Growth of these plates is not reversible on heating or removal/reversal of an applied stress because this plastic accommodation strain is irreversible.

However, in certain ferrous alloys there is a reduction in the elastic modulus sufficient to permit growth to be accommodated elastically, when plates form in groups or clusters of crystallographic variants that permit self-accommodation of the shear component of the shape strain. Such assemblies of plates are in a thermoelastic or mechanoelastic equilibrium, so that they form and grow when the temperature is lowered below M_s or a shear stress is applied immediately above M_s, and shrink and disappear if the sign of the chemical or mechanical driving force is reversed. Thermoelastic martensites have been observed in, for example, Fe–Pt, Fe–Pd, and Fe–Ni–Co–Ti alloys.

Global transformation kinetics, which in essence define the volume fraction of transformed product as a function of time and temperature, may be either athermal or isothermal. In the former, the transformation proceeds only with decreasing temperature, while in the latter the transformation progresses with time at a constant temperature and exhibits classical C-curve behavior, even though the temperature may be well below the range in which significant diffusion may occur. For both modes of transformation, once initiated the transformation is rarely continuous but tends to proceed in discontinuous "bursts" with decreasing temperature (athermal transformation) (Fig. 1), or increasing time at temperature (isothermal transformation). Each burst is the product of autocatalytic nucleation and rapid growth of numerous plates. Individual plates grow to near final size very rapidly, so that transformation proceeds predominantly via the nucleation and growth of new plates. Transformations in most ferrous alloys exhibit athermal transformation kinetics; however, there are systems that exhibit isothermal martensitic transformations, including, for example, certain Fe–Ni–Mn alloys where transformation occurs at temperatures well below ambient.

Bibliography

Cahn J W 1977 The symmetry of martensites. *Acta Metall.* **25**, 721–4

Christian J W 1975 *Theory of Transformations in Metals and Alloys*, 1st edn Pergamon, Oxford

Delaey L 1991 Diffusionless transformations. In: Cahn R W, Haasen P, Kramer E J (eds.) *Materials Science and Technology, Vol. 5, Phase Transformations in Materials.* VCH, Weinheim, Germany

Honeycombe R W K, Bhadeshia H K D H 1995 *Steels, Microstructure and Properties*, 2nd edn. Arnold, London

Muddle B C 1982 The crystallography and morphology of ferrous martensites. In: Aaronson H I, Laughlin D E, Sekerka R F, Wayman C M (eds.) *Proc. 1st Int. Conf. on Solid→Solid Phase Transformations.* TMS-AIME, Warrendale, PA

Nishiyama Z 1978 *Martensitic Transformation.* Academic Press, New York

Olson G B, Owen W S (eds.) 1992 *Martensite.* ASM International, Materials Park, OH

Wayman C M 1964 *Introduction to the Crystallography of Martensitic Transformations.* Macmillan, New York

B. C. Muddle and J. F. Nie
Monash University, Clayton, Victoria, Australia

N

Nanoscale Ceramic Composites

A new class of structural ceramic materials with superior mechanical properties is currently being developed. These materials are known as "ceramic nanocomposites" and are produced by combining a structural ceramic matrix, such as Al_2O_3, ZrO_2, or Si_3N_4 with a nano-sized second phase. The second phases typically used are "nanoparticles" which are 300 nm or less in size and added in quantities between 1 vol.% and 30 vol.% (Todd 1995). Ceramic nanocomposites have been reported to possess both room- and high-temperature properties which are greatly superior to those of the monolithic matrix in a variety of material systems, indicating that a "structural synergism" occurs between the matrix and second phase in these systems (Niihara and Nakahira 1988, Ding *et al.* 1993, Pezzotti and Sakai 1994). In this article, the development of the most widely studied ceramic nanocomposite, Al_2O_3–SiC, will be discussed. In particular, the processing steps required to fabricate Al_2O_3–SiC nanocomposites and a review of their microstructures will be given. Most importantly, the improvements in mechanical properties over monolithic Al_2O_3 will be presented, and the possible mechanisms responsible for these improvements will be discussed.

1. Processing and Microstructure of Al_2O_3–SiC Nanocomposites

Al_2O_3–SiC nanocomposites are produced using standard ceramic processing techniques. Matrix and second-phase powders are typically comilled in an organic medium to achieve a uniform dispersion of second phase throughout the matrix. After milling, the nanocomposite powder is dried and then densified using pressureless sintering or hot pressing. Relative to that of monolithic Al_2O_3, densification in the nanocomposite is strongly inhibited by the presence of particles on grain boundaries, which reduce the rate of grain-boundary diffusion. As a result, the sintering conditions necessary to achieve full densification are strongly dependent on the amount of second phase, with the sintering temperature increasing significantly with the second-phase volume fraction. While monolithic Al_2O_3 can be pressureless sintered to near-theoretical density at 1400 °C, a temperature of 1775 °C is necessary for full densification in a nanocomposite of Al_2O_3 containing 5 vol.% SiC (Stearns *et al.* 1992). In fact, no study has shown that full density can be achieved using pressureless sintering in a nanocomposite containing more than 5 vol.% SiC. The difficulties of densification in the nanocomposite can be alleviated by hot pressing. During hot pressing, an applied pressure increases the driving force for densification, allowing dense nanocomposites to be manufactured at lower temperatures (e.g., 1500 °C for 5 vol.% SiC) and with higher volume fractions of second phase (e.g., 20 vol.% SiC at a temperature of 1675 °C) (Stearns and Harmer 1996).

The processing steps used to produce nanocomposites have a significant effect on their microstructures and properties. It is well known that the properties of the nanocomposite are strongly dependent on the second-phase dispersion (Stearns *et al.* 1992). Specifically, a uniform dispersion of second phase is critical to obtaining superior properties in the nanocomposite relative to the monolith. Unfortunately, the spatial distribution of the second phase has been shown to be very sensitive to the processing conditions used, indicating that small changes in processing can lead to significant differences in the properties of the nanocomposite.

The microstructures of nanocomposites produced using the above technique are also strongly dependent on the amount of second phase present. In nanocomposites with a high volume fraction of second phase (>5 vol.%), matrix grains are equiaxed with a narrow size distribution, and second-phase particles are uniformly distributed among the matrix grains. An example of this type of microstructure (5 vol.% SiC in Al_2O_3) can be seen in Fig. 1. On the contrary, in samples with a low volume fraction of second phase, the grain-size distribution of matrix grains may be significantly wider and in extreme cases, abnormal grains may be present.

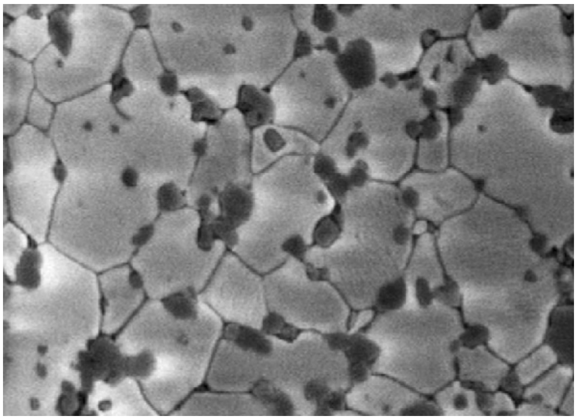


Figure 1
Scanning-electron-microscope image of the microstructure of a Al_2O_3–5 vol.% SiC nanocomposite showing the presence of both intergranular and intragranular particles.

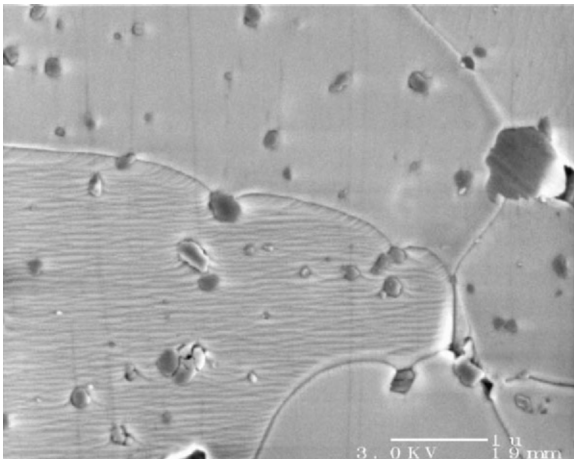


Figure 2
Scanning-electron-microscope image of SiC particles (5 vol.%) inhibiting grain growth by pinning migrating Al_2O_3 grain boundaries.

As shown in Fig. 1, second-phase particles in the nanocomposite are either located on grain boundaries (intergranular) or inside grains (intragranular). As will be discussed later, the location of particles with respect to grain boundaries plays the paramount role in the improvement of the properties of the nanocomposite over the monolith (Xu *et al.* 1997). In particular, intragranular particles are believed to be responsible for the superior room-temperature properties of the nanocomposite, while the intergranular particles are believed to be responsible for its superior high-temperature properties. Intergranular particles also significantly increase the microstructural stability of the nanocomposite. The interaction of particles with grain boundaries strongly inhibits matrix grain growth which would normally take place at high temperatures. Figure 2 shows SiC particles inhibiting grain growth in the nanocomposite by pinning migrating grain boundaries. The relative proportions of intra- and intergranular particles is strongly related to the matrix grain size in the Al_2O_3–SiC system, with the percentage of intergranular particles decreasing significantly with the grain size. For example, approximately 70% of particles are expected to be intergranular for a grain size of 1 μm, whereas 25% of particles are expected to be intergranular for a grain size of 5 μm (Stearns and Harmer 1996).

2. Room-temperature Properties of Al_2O_3–SiC Nanocomposites

The superior mechanical properties of Al_2O_3–SiC nanocomposites at room temperature were first shown by Niihara, who found that the addition of 5 vol.% of 300 nm SiC particles to Al_2O_3 increased the strength from 380 MPa to 1100 MPa (Niihara and Nakahira 1988). Furthermore, it was shown that a simple annealing treatment at 1300 °C for 1 hour in argon gas subsequently increased the strength of the nanocomposites to over 1500 MPa. Niihara reported that the addition of the SiC particles also resulted in a significant increase in the fracture toughness, from 3.8 MPa m$^{1/2}$ to 4.7 MPa m$^{1/2}$. The superior properties of Al_2O_3–SiC nanocomposites over monolithic Al_2O_3 have been confirmed by a number of researchers, although none of these improvements have been as dramatic as those reported by Niihara (Zhao *et al.* 1993, Borsa *et al.* 1995, Jang *et al.* 1996).

While the superior properties of the nanocomposite are well documented, the mechanisms responsible for the improvement in properties are still unclear. Proposed strengthening and toughening mechanisms for the unannealed nanocomposite include the reduction in critical flaw size due to the development of intragranular particle-induced subgrain boundaries, the attraction of intergranular cracks into matrix grains by residual tensile hoop stresses which arise due to the large thermal-expansion mismatch between Al_2O_3 and SiC, and the formation of a surface layer of compressive stresses during machining. The subsequent increase in strength and toughness after annealing has been attributed to the healing of surface flaws and the retainment of machining-induced compressive surface stresses, both the result of a slower stress relaxation behavior in the nanocomposite (Thompson *et al.* 1995).

In addition to having improved strength and toughness relative to monolithic Al_2O_3, Al_2O_3–SiC nanocomposites also have superior wear properties. The improved wear properties are the result of a change in fracture mode which occurs when the concentration of the SiC second phase is greater than ~4 vol.%. While monolithic Al_2O_3 exhibits predominantly intergranular fracture, the large thermal-expansion mismatch between Al_2O_3 and SiC results in an apparent strengthening of Al_2O_3 grain boundaries, causing the nanocomposite to fracture transgranularly. As a result, grains in Al_2O_3–SiC nanocomposites are worn away slowly, in contrast to monolithic Al_2O_3 where wear typically involves the removal of entire grains due to the relatively weak interface at grain boundaries. The effect of the change in fracture mode on the wear properties of the nanocomposite is dramatic. In one study, the wet erosive wear rate of a 5 vol.% SiC nanocomposite was measured to be one-third that of monolithic Al_2O_3 with the same grain size (Walker *et al.* 1994). Furthermore, the lack of grain pullout in the nanocomposite allows a higher-quality surface finish to be prepared (and retained during wear) than is possible with monolithic Al_2O_3.

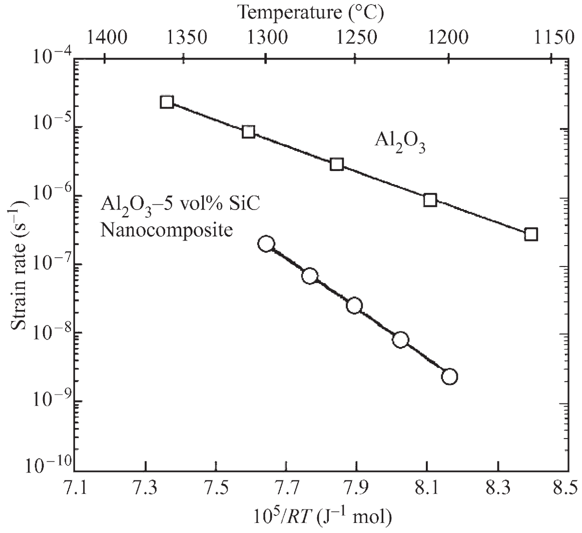


Figure 3
Plot of creep strain rate as a function of temperature for monolithic Al_2O_3 and Al_2O_3–5 vol.% SiC nanocomposites (after Thompson *et al.* 1997).

3. High-temperature Properties of Al_2O_3–SiC Nanocomposites

In addition to its superior room-temperature properties, Al_2O_3–SiC nanocomposite also possesses superior high-temperature properties. The superior strength and toughness of the nanocomposite at room temperature are retained as high as 1000 °C (Niihara and Nakahira 1988). Furthermore, studies have shown that at higher temperatures, the creep resistance of Al_2O_3 is increased significantly by the presence of SiC particles (Ohji *et al.* 1995, Thompson *et al.* 1997). As shown in Fig. 3, the creep rate of Al_2O_3 containing 5 vol.% SiC (150 nm) is 2–3 orders of magnitude lower than that of monolithic Al_2O_3 at temperatures as high as 1300 °C. In addition, the creep life of the nanocomposite is more than an order of magnitude longer than in the monolith. Unfortunately, the failure strain of the nanocomposite is lower than that of Al_2O_3 due to the nucleation of cavities at intergranular particles. The improved creep resistance of Al_2O_3–SiC has been attributed to a number of mechanisms. These mechanisms include a reduction in the rate of grain-boundary diffusion due to the presence of intergranular particles; an inhibition of shape-accommodation processes necessary for creep to occur, such as grain-boundary sliding and migration; and a reduction of the intrinsic grain-boundary diffusivity of Al_2O_3 by silicon cation contamination. The dominant mechanism responsible for the improved creep behavior of the nanocomposite is unknown.

4. Concluding Remarks

Ceramic nanocomposites are relatively new materials which appear to have promise. In particular, Al_2O_3 containing nano-sized SiC second-phase particles has been shown to possess both room- and high-temperature properties which are superior to monolithic Al_2O_3. These properties, which include higher strength and toughness, increased wear resistance, lower creep rates at elevated temperatures, and excellent microstructural stability, make nanocomposites an attractive alternative to monolithic ceramics for many structural applications.

Bibliography

Borsa C E, Jiao S, Todd R I, Brook R J 1995 Processing and properties of Al_2O_3/SiC nanocomposites. *J. Microsc.* **177** (3), 305–12

Ding Z, Oberacker R, Thümmler F 1993 Microstructure and mechanical properties of yttria-stabilized tetragonal zirconia polycrystals (Y-TZP) containing dispersed silicon carbide particles. *J. Eur. Ceram. Soc.* **12** (5), 377–83

Jang B-K, Enoki M, Kishi T, Lee S -H, Oh H-K 1996 Fracture behavior and toughening of alumina-based composites fabricated by microstructural control. In: Bradt R C, Munz D, Sakai M, Shevchenko V Ya (eds.) *Fracture Mechanics of Ceramics*. Plenum, New York, Vol. 12, pp. 371–82

Jiao S, Jenkins M L, Davidge R W 1997 Electron microscopy of crack/particle interactions in Al_2O_3/SiC nanocomposites. *J. Microsc.* **185** (2), 259–64

Niihara K 1991 New design concept of structural ceramics: ceramic nanocomposites. *The Centennial Memorial Issue of The Ceramic Society of Japan* **99** (10), 974–82

Niihara K, Nakahira A 1988 Strengthening of oxide ceramics by SiC and Si_3N_4 dispersions. *Proc. 3rd Int. Symp. Ceramic Materials and Components for Engines*. The American Ceramic Society, Westerville, OH, 919–26

Niihara K, Nakahira A 1990 Particulate-strengthened oxide nanocomposites. In: Vincenzini P (ed.) *Advanced Structural Inorganic Composites*. Elsevier, London, pp. 637–64

Niihara K, Nakahira A, Sasaki G, Hirabayashi M 1989 Development of strong Al_2O_3–SiC composites. *Materials Research Society Symp. Proc. Int. Meeting on Advanced Materials*. Materials Research Society, Warrendale, PA, Vol. 4, pp. 129–34

Ohji T, Nakahira A, Hirano T, Niihara K 1995 Tensile creep behavior of alumina/silicon carbide nanocomposite. *J. Am. Ceram. Soc.* **77** (12), 3259–63

Pezzotti G, Sakai M 1994 Effect of a silicon carbide "nano-dispersion" on the mechanical properties of silicon nitride. *J. Am. Ceram. Soc.* **77** (11), 3039–41

Stearns L C, Harmer M P 1996 Particle-inhibited grain growth in Al_2O_3–SiC: I, experimental results. *J. Am. Ceram. Soc.* **79** (12), 3013–9

Stearns L C, Zhao J, Harmer M P 1992 Processing and microstructure development in Al_2O_3–SiC "nanocomposites." *J. Eur. Ceram. Soc.* **10** (6), 473–7

Thompson A M, Chan H M, Harmer M P 1995 Crack healing and stress relaxation in Al_2O_3–SiC "nanocomposites."*J. Am. Ceram. Soc.* **78** (3), 567–71

Thompson A M, Chan H M, Harmer M P 1997 Tensile creep of alumina–silicon carbide "nanocomposites." *J. Am. Ceram. Soc.* **80** (9), 2221–8

Todd R I 1995 Alumina/SiC nanocomposites: big things from small particles. *Br. Ceram. Proc.* **55**, 225–38

Walker C N, Borsa C E, Todd R I, Davidge R W, Brook R J 1994 Fabrication, characterisation and properties of alumina-matrix nanocomposites. *Br. Ceram. Proc.* **53**, 249–64

Xu Y, Zangvil A, Kerber A 1997 SiC nanoparticle-reinforced Al_2O_3 matrix composites: role of intra- and intergranular particles. *J. Eur. Ceram. Soc.* **17**, 921–8

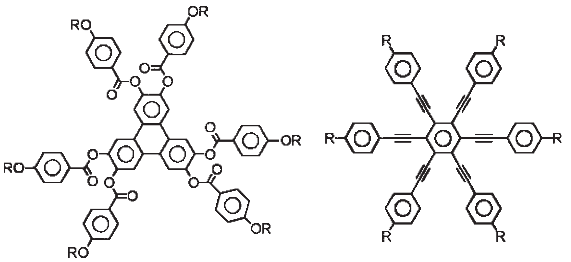
Zhao J, Stearns L C, Harmer M P, Chan H M, Miller G A 1993 Mechanical behavior of alumina-silicon carbide "nanocomposites." *J. Am. Ceram. Soc.* **76** (2), 503–10

G. S. Thompson and M. P. Harmer
Lehigh University, Bethlehem, Pennsylvania, USA

Nematic Discotic Liquid Crystals

The discotic nematic (N_D) phase consists of an orientational ordered arrangement of disk-like molecules but with no long-range transitional order. The director defining the preferred axis of orientation of the molecules is the molecular short axis or the disk normal.

The first example of an N_D phase was observed in a triphenylene derivative (Fig. 1(a)). With $R = O\text{-}C_8H_{17}$ the N_D phase occurs in the temperature range between 180 and 222 °C. Although a large number of discotic liquid crystals have been synthesized in the meantime, most of them show higher ordered columnar phases rather then the nematic phase. Thus, only a comparatively small number of discotic nematic liquid crystals are known so far, and a more or less complete set of physical properties is available for only a few.



(a) (b)

Figure 1
Two examples of nematic discotic liquid crystals. (a) First nematic discotic liquid crystal, a triphenylene derivative. (b) Principal structure of the model compounds, based on a benzene-centered multiyne core.

Therefore, most of the results presented here refer to a special class of discotic liquid crystals as model systems (Fig. 1(b)) (Chandrasekhar 1998).

1. *Properties of the Discotic Nematic Phase N_D*

For the discotic nematic phase as for the calamitic one the anisotropy of various physical properties, such as refractive index (n), dielectric permittivity (ε), and magnetic susceptibility (χ) is related to the order parameter $S = \frac{1}{2}\langle 3\cos^2\theta - 1 \rangle$.

Consequently, the temperature dependence of the anisotropic values is mainly owing to the temperature dependence of S, whereas the magnitude as well as the sign of the anisotropy is determined by the molecular structure.

In contrast to almost all rod-like molecules, the flat molecular structure causes a negative sign of both the optical and the magnetic anisotropy (Fig. 2). These are characteristic features of all discotic liquid crystals. As a consequence it is not possible to achieve an uniaxial alignment by applying a homogeneous magnetic field (Levelut *et al.* 1981).

For the dielectric anisotropy both signs are found, depending on the presence of permanent dipole moments and their orientation with respect to the molecular short axis, i.e., the disk-normal. Since the selected compounds do not possess any dipole moment, the dielectric anisotropy is also negative in this case (Fig. 2).

Generally, the order parameter, S, is directly proportional to the anisotropy of the magnetic mass

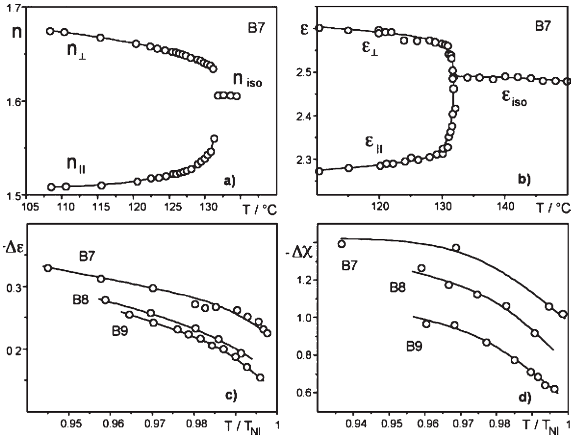


Figure 2
Temperature dependence of anisotropic properties of the discotic nematic phase (multiyne derivatives, where R = heptyl (B7), octyl (B8), and nonyl (B9), see Fig. 1(b)): (a) refractive indices, (b) dielectric constants, (c) dielectric anisotropy, (d) magnetic susceptibility anisotropy (10^{-6} SI units) (after Sabaschus 1992, reproduced with permission).

susceptibility, which for the discotic case has been determined only for very few liquid crystals. Neglecting small corrections for the inner field, the order parameter, S, can also be easily determined from the optical anisotropy, which in the absence of permanent dipole moments is proportional to the dielectric anisotropy. Using the same standard extrapolation procedure as for calamitic liquid crystals, absolute values of S can be obtained.

As Fig. 3 shows, all three different physical properties give the same temperature dependence for the naphthalene-centered multiyne under investigation.

At low temperatures the experimental values agree with the order parameter calculated by the Maier–Saupe theory, but while approaching the clearing temperature, distinct deviations are observed.

Quite probably this effect is owing to an increased twist of the outer benzene rings at higher temperature reducing the molecular anisotropy mainly caused by the aromatic rings. In conventional rod-like molecules such a twisting will not affect the molecular anisotropy because of the molecular rotation around the long axis.

A characteristic property of the nematic phase is its elastic behavior favoring a uniform director orientation. There are three basic deformations: splay, twist, and bend, which are described by the elastic constants K_1, K_2, and K_3 respectively.

The elastic constants of the discotic nematic phase have been measured only for a few compounds (Phillips *et al.* 1993, Chandrasekhar 1998). For some multiyne compounds K_3 and K_1 could be obtained by the electric field induced deformation and K_2 from the critical cell-thickness for which a change from a helical structure to a uniform orientation takes place in chiral doped samples (Fig. 4).

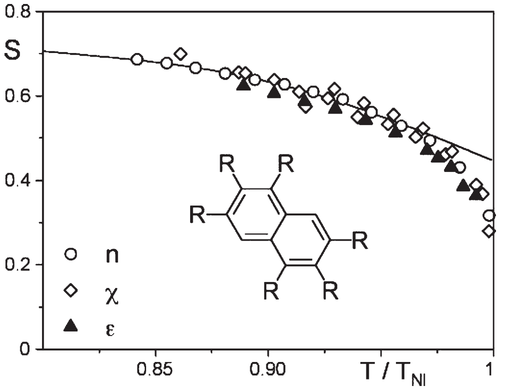


Figure 3
Temperature dependence of the order parameter obtained from the dielectric (ε), magnetic (χ), and optical (n) anisotropies in comparison to the order parameter predicted by the Maier–Saupe theory (solid line). Naphthalene-centered multiyne with $R = -C\equiv C-\varnothing-C_9H_{19}$ (after Sabaschus 1992, reproduced with permission).

It is interesting to note that the values are in the same order of magnitude as that obtained for calamitic liquid crystals. Considering ellipsoids with different shape anisotropy, a reversed sequence of the elastic constants is predicted for disk-like molecules in comparison to rod-like nematic liquid crystals: $K_2 > K_1 > K_3$ (Sokalski and Ruijgrok 1982), being confirmed by the experimental investigations (Fig. 5).

As for calamitic liquid crystals the temperature dependence of the elastic constants is approximately given by the square of the order parameter ($K_i \sim S^2$).

For a complete description of the viscoelastic properties, the set of five viscosity coefficients is needed, which have not yet been determined for a discotic nematic liquid crystal. Only a few values of effective viscosity parameters were measured. At the same temperature the viscosity exceeds that of calamitic liquid crystals by two orders of magnitude. Concerning flow alignment, it was predicted that owing to the disk-like shape of the molecules the director will adopt negative angles with respect to the flow direction, not positive angles as in the calamitic case (Chandrasekhar 1998).

2. Physical Properties of the Discotic Cholesteric Phase

It is well known for calamitic systems, that chiral molecular structure present either in the liquid crystal compound or in a dopant will twist the nematic to the chiral nematic phase (cholesteric phase). A similar helical structure is exhibited by chiral twisting a nematic phase consisting of disk-like molecules as shown in Fig. 5. The preferred orientation of the short axis of the disks, i.e., the director ***n*** rotates

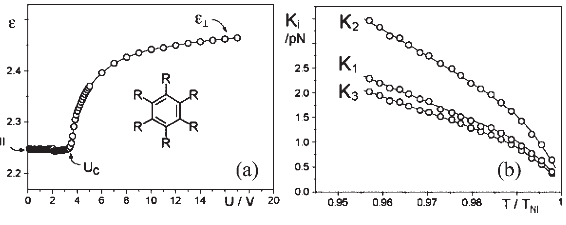


Figure 4
(a) Apparent dielectric constant of an initially homeotropically aligned sample as a function of the applied voltage. From the threshold voltage U_C the bend elastic constant K_3 is calculated, whereas K_1 is deduced from fitting the rise of the apparent dielectric constant. K_2 was obtained from the critical cell thickness with chiral doped systems in which a transition from twisted to the untwisted state takes place (see Fig. 6). (b) Elastic constants of the multiyne compound B9 (see Fig. 1(b)) (after Krüerke 1999).

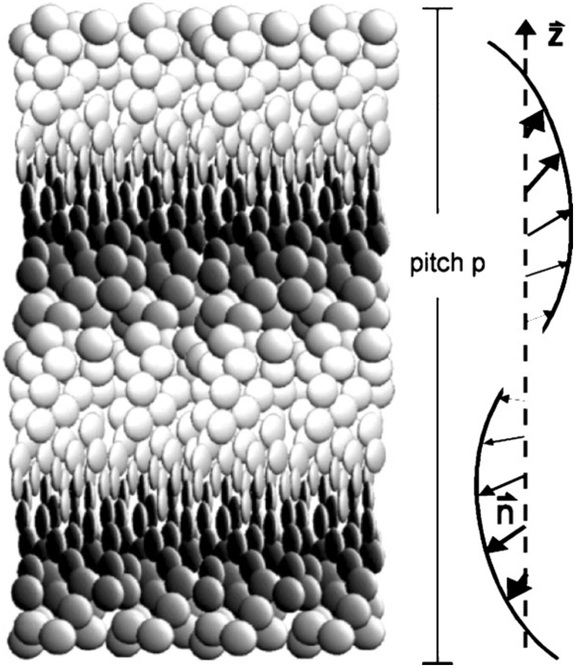


Figure 5
Schematic representation of the discotic cholesteric phase. The local director ***n*** is propagating around the axis *z* describing a left-handed or right-handed helical structure of a certain pitch *p*.

about the helix-axis (*z*). The pitch, *p*, is the distance of one full revolution. Right-handed ($p>0$) and left-handed ($p<0$) helical structures can be formed.

Replacing the nonchiral side chains of the nematic triphenylene (see Fig. 1) by chiral substituents lead to the first example of a discotic cholesteric phase ($R=O\text{-}(CH_2)_3CH(CH_3)C_2H_5$)) with a pitch of $\sim 30\,\mu m$ at 200 °C (Chandrasekhar 1998). As for calamitic liquid crystals the helical pitch can be determined either from the periodicity appearing in a fingerprint texture in samples prepared with the helical axis parallel to the substrates or by the distance between cano-lines in a wedge-shape sample with the helical axis normal to the substrates. Since discotic liquid crystals tend to adopt a homeotropic alignment, the Cano method generally cannot be applied. Quite remarkably for multiyne compounds, weak anchoring is often observed. Thus, cholesteric discotic phases based on multiyne compounds exhibit the grandjean texture with cano-lines at larger thicknesses of the wedge, whereas at smaller thicknesses a transition to the homeotropic alignment takes place and a fingerprint texture is observed (Fig. 6). At even smaller sample thickness a transition to the untwisted homeotropic aligned nematic phase occurs, allowing determination of the twist-elastic constant, K_2, from the critical thickness $d_0=p/2\cdot K_3/K_2$ (Fischer 1976).

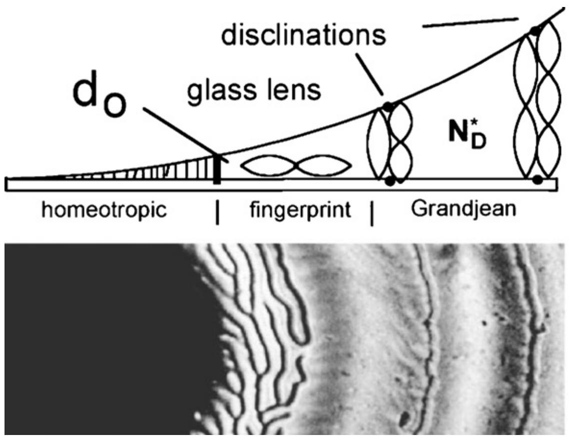


Figure 6
Schematic representation of the Cano experiment using planar alignment conditions with a discotic cholesteric liquid crystal exhibiting weak anchoring and corresponding view through the polarizing microscope.

Using discotic multiyne compounds it was demonstrated that as in the case of calamitic liquid crystals depending on the lateral substituents, either a short helical pitch or a temperature-induced helix inversion can be realized. With the S-enantiomer of a radial pentayne, with $R=OC_2H_4CH(CH_3)C_3H_6CH(CH_3)_2$, $X=OC_{16}H_{33}$ (see Fig. 7), selective reflection of right-handed circularly polarized light is observed in the temperature range between −25 °C and 10 °C. Taking into account the mean refractive index ($n=1.73$) the value of the pitch could be calculated from the wavelength of maximum reflection to be $p=0.3\ \ldots\ 0.4\,\mu m$. With a substituent $R=OCH_2CH(CH_3)C_2H_5$, a cholesteric phase is obtained which shows an unusual temperature dependence of the pitch resulting in a helix inversion at about 97 °C (Fig. 7). Dissolving the chiral compound in a nematic induced multiyne cholesteric phases. The pitch induced is inversely proportional to the concentration of the dopant (*c*). The constant of proportionality, which depends both as the nematic host and the temperature, is called the helical twisting power of the chiral dopant. As Fig. 7 shows, the *pc* product is found to be independent of composition, thereby establishing the relation $1/p\sim c$ (chiral dopant), which is well known for induced cholesteric phases with calamitic liquid crystals. Remarkably, the inversion temperature does not depend on the concentration, showing that the inversion effect is an intrinsic property of the chiral dopant (Langner *et al.* 1995).

Obviously, with suitable molecular structures, values of pitch and temperature dependencies similar to those of calamitic cholesteric phases can be produced. However, much less is known about the relationship

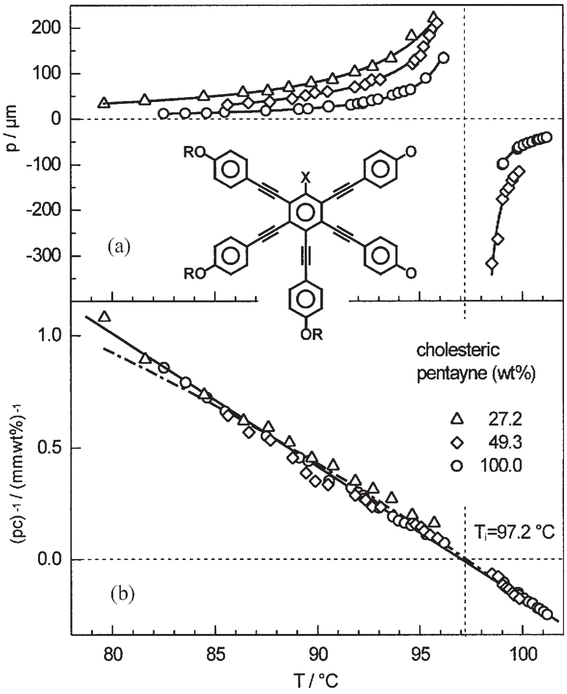


Figure 7
Temperature dependence of the helical pitch (a), and the helical twisting power (b) for a pure discotic cholesteric pentayne ($R = CH_2CH(CH_3)CH_2CH_3$, $X = OC_{16}H_{33}$) which exhibits a temperature-induced helix inversion at $T_i \sim 97\,°C$, and for diluted mixtures with a suitable discotic nematic compound (Langner *et al.* 1995, reproduced by permission of from *Journal of Materials Chemistry*, 1995, **5**, 693–9).

between the chiral molecular structure and the helix formation.

It is interesting to note that extremely high helical twisting powers were found with chiral compounds such as sugar derivatives (Krüerke *et al.* 1993) and nonliquid crystalline chiral substituted hexaynes (Booth *et al.* 1996). These compounds induce short-pitch cholesterics ($p < 0.3\,\mu m$) with concentrations below 20 wt.%, although they do not possess a cholesteric phase themselves.

As in calamitic cholesteric systems with high chirality, up to three blue phases BP_DI, BP_DII, BP_DIII were observed with these induced cholesteric discotic systems. Single crystals of BP_DI and BP_DII can also be grown (Fig. 8). A chiral triphenylene derivative has been synthesized, which is the first pure discotic compound exhibiting a Blue Phase (Heppke *et al.* 2000).

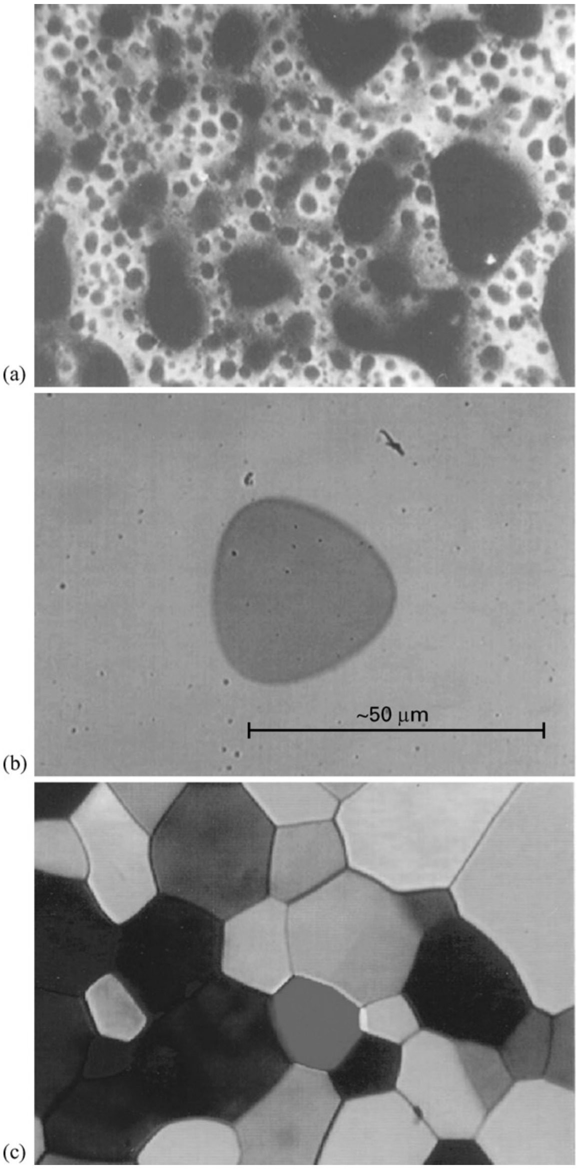


Figure 8
Textures of discotic blue phases observed under a polarizing microscope on cooling from the isotropic liquid: (a) BP_DIII or "blue fog", (b) BP_DII single crystal, (c) BP_DI platelet texture.

A special feature of discotic compounds in general is the pronounced tendency to form a vitrified state, which enables the preservation of liquid crystalline order in the glassy state with moderate cooling rates. Exploiting this, it was possible to use the special method of Kossel spectroscopy to determine the symmetry and lattice constants of both BP_DI and BP_DII, which turned out to be the direct analogues of the corresponding blue phases formed by calamitic cholesteric systems. Neither the critical pitch at which the blue phase occurs, nor the structure of blue

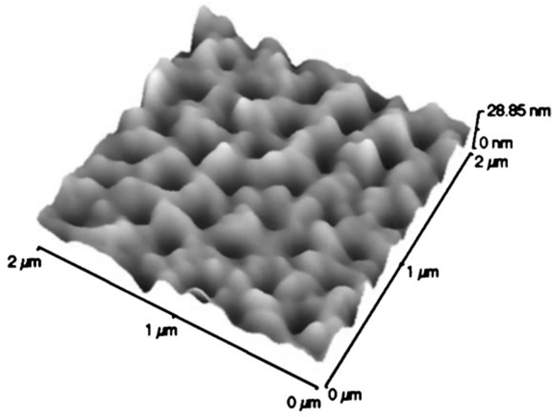


Figure 9
Free surface image of a frozen BP_DI structure obtained by atomic force microscopy (Hauser *et al.* 1997, reproduced by permission of from *Journal of Materials Chemistry*, 1997, **7**, 2223–9).

phases seem to be significantly altered by the different shape anisotropy of discotic liquid crystals.

Also, the free surface of vitrified discotic blue phases could be investigated for the first time by atomic force microscopy (Fig. 9). The surface is strongly modulated at a periodicity which corresponds to the lattice constants of the respective blue phases, but no indication of defects owing to disclination lines could be found (Hauser *et al.* 1997).

3. Special Properties and Potential Applications

As demonstrated so far, there are no fundamental differences between nematic and cholesteric phases of rod- and disk-like molecules. Essentially, the director seems not to care about the molecular shape anisotropy. The peculiar phenomenon of an inverse or even re-entrant phase sequence N_D-Col(-N_D) has been reported (Chandrasekhar 1998), but a theoretical analysis is still lacking. Some electro-optical effects such as dynamic scattering (Dubois *et al.* 1981), Frederiks field effect (Heppke *et al.* 1988) and an electrically tunable cholesteric mirror (Krüerke *et al.* 2000a) were also demonstrated but never made it to application.

However, biaxiality is especially related to molecular shape. Nematic biaxiality is anticipated to appear between the calamitic and the discotic nematic phases. Although numerous attempts have been made to observe it, there is still a doubt about its existence for low molecular weight thermotropic systems.

Owing to their molecular shape, discotic compounds exhibit negative birefringence ($\Delta n < 0$). This property is commercially used in polymer compensation films to enhance the color characteristics of wide viewing angle STN (super-twist nematic) display devices (Mori *et al.* 1997).

Potential nondisplay applications have also been demonstrated: the high tendency to form a glassy state makes discotic liquid crystals suitable materials for holographic information storage at reasonable temperatures (Contzen *et al.* 1996). The special class of multiyne compounds shows a highly sensitive intrinsic photoelectric reorientation effect (MacDonald *et al.* 1999). The use of discotic materials as lubricants has been demonstrated (Eidenschink *et al.* 1999).

A special kind of nematic phase is obtained by mixing discotic materials with electron acceptors such as trinitrofluorenone: charge transfer interaction leads to the formation of columnar rod-like supramolecular assemblies, thereby inducing the so-called calamitic nematic phase, N_C, in a certain range of concentration (Chandrasekhar 1998). Similar nematic and cholesteric phases are obtained by dilution of columnar liquid phases with suitable hydrocarbons such as dodecane. These systems exhibit phenomena such as selective reflection and polar switching (Krüerke 2000b).

Bibliography

Booth C J, Krüerke D, Heppke G 1996 Highly twisting enantiomeric radial multiyne dopants for discotic liquid-crystalline systems. *J. Mater. Chem.* **6** (6), 927–34

Chandrasekhar S 1998 In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals: Low Molecular Weight Liquid Crystals II*. Vol. 2B, Wiley-VCH, Weinheim, Germany.

Contzen J, Heppke G, Kitzerow H-S, Krüerke D, Schmid H 1996 Storage of laser-induced holographic gratings in discotic liquid crystals. *Appl. Phys. B* **63**, 605–8

Dubois J C, Hareng M, Le Berre S, Perbet J N 1981 Dielectric and hydrodynamic instabilities in certain classes of discotic mesophases. *J. Phys.* **38** (1), 11–3

Eidenschink R, Konrath G, Kretzschmann H, Rombach M 1999 Ununsual 1999 Lift by shearing mesogenic fluids. *Mol. Cryst. Liq. Cryst.* **330**, 327–34

Fischer F 1976 Critical pitch in thin cholesteric films with homeotropic boundaries. *Z. Naturforsch. A* **31**, 41–6

Hauser A, Thieme M, Saupe A, Krüerke D, Heppke G 1997 Surface imaging of frozen blue phases in a discotic liquid crystal with atomic force microscopy. *J. Mater. Chem.* **7** (11), 2223–9

Heppke G, Kitzerow H, Oestreicher F, Quentel S, Ranft A 1988 Electrooptic effect in a non-polar nematic discotic liquid crystal. *Mol. Cryst. Liq. Cryst.* **6** (3), 71–9

Heppke G, Krüerke D, Löhning C, Lötzsch D, Moro D, Müller M, Sawade H 2000 New chiral discotic triphenylene derivatives exhibiting a cholesteric blue phase and a ferroelectrically switchable columnar mesophase. *J. Mater. Chem.* **10**, 2657–61

Krüerke D 1999 *Experimentelle Untersuchungen nematischer und cholesterischer Phasen Diskotischer Flüssigkristalle*. Thesis TU-Berlin, Mensch & Buch, Berlin, Germany

Krüerke D, Gough N, Heppke G, Lagerwall S T 2000a Electrically tunable cholesteric mirror. *Mol. Cryst. Liq. Cryst.* **351**, 69–78

Krüerke D, Kitzerow H-S, Heppke G, Vill V 1993 First observation of selective reflection and blue phases in chiral discotic liquid crystals. *Ber. Bunsenges. Phys. Chem.* **97**, 1371–5

Krüerke D, Rudquist P, Lagerwall S T, Sawade H, Heppke G 2000b Ferroelectric switching of chiral discotic lyomesophases. *Ferroelectrics* **243**, 207–20

Langner M, Praefcke K, Heppke G, Krüerke D 1995 Chiral radial pentaynes exhibiting cholesteric discotic phases. *J. Mater. Chem.* **5** (4), 693–9

Levelut A M, Hardouin F, Gasparoux H, Destrade C, Tinh N H 1981 X-ray investigations and magnetic field effect on a nematic phase of disc-like molecules. *J. Phys.* **42**, 147–52

Macdonald R, Meindl P, Busch S 1999 The photoelectrical reorientation effect in nematic liquid crystals. *J. Nonlin. Opt. Phys. Mater.* **8** (3), 379–89

Mori H, Itoh Y, Nishiura Y, Nakamura T, Shingawa Y 1997 Performance of a novel optical compensation film based on negative birefingence of discotic compounds for wide-viewing-angle twist-nematic liquid crystal displays. *Jpn. J. Appl. Phys.* **36**, 143

Phillips T J, Jones J C, McDonnel D G 1993 On the influence of short range order upon the physical properties of triphenyline nematic discogens. *Liq. Cryst.* **15** (2), 203–15

Sabaschus B 1992 *Physikalische Eigenschaften diskotisch nematischer Phasen.* Thesis TU, Berlin, Germany

Sadashiva B K 1998 In: Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) *Handbook of Liquid Crystals: Low Molecular Weight Liquid Crystals II.* Vol. 2B, Wiley-VCH Weinheim, Germany.

Sokalski K, Ruijgrok Th W 1982 Elastic constants for liquid crystals of disc-like molecules. *Physica A* **113**, 126–32

G. Heppke and D. Krüerke
Technische Universität Berlin, Germany

Nickel Alloys: Nomenclature

The numbering and identification of a nickel alloy is initially defined by the individual producers and users of the alloy. This has led to a profusion of trademarks and specific alloy designations. The Special Metals Corporation sells product under the trademarks BRIGHTRAY, INCONEL, INCOLOY, MONEL, NIMONIC, and UDIMET, among others. Haynes International uses the trademarks HASTELLOY and HAYNES. Other trademarks used for nickel alloys and nickel-based superalloys include MAR-M by Martin Metals Corporation, Rene by General Electric Corporation, PWA by Pratt & Whitney Aircraft Division, United Technologies Corporation, and Nicrofer by Krupp VDM GmbH. Examples of specific alloy designations include RENE 41 by Teledyne Allvac, alloy 625LCF by Special Metals Corporation, and CMSX-4 by Cannon–Muskegon Corporation. In many instances, different trademarks are used for the same alloy produced by the several manufacturers.

There is presently no worldwide clearing-house dedicated specifically to nickel alloys. Without this specific unified system of alloy identification, the potential exists for confusion due to more than one identification number being used for the same alloy, or for the same number to be used for two or more different alloys. This gap is filled by generic coding and identification systems: for example, the International Organization for Standardization (ISO), the Unified Numbering System (UNS), and the Werkstoff system. Table 1 lists examples of common names of alloys and the appropriate classification numbers.

The ISO publishes a list of designations for nickel alloys. It is an international agency for standardization comprising the national standards entities of about 90 countries. The interests of users, producers, governments, and the scientific community are brought together and the international exchange of goods and services is facilitated. Use of these designations simplifies communication related to alloys.

The Unified Numbering System for Metals and Alloys (UNS) correlates many metal and alloying identification systems administered by societies and trade associations. Assignments for the alloys are processed by the American Society for Testing and Materials, West Conshohocken, PA, and the Society of Automotive Engineers, Warrendale, PA. The UNS number is only an identifier of an alloy for which limits have been established in specifications published elsewhere. The UNS designation is not a specification as it defines no requirements for form, condition, or properties.

Table 1
Examples of nickel alloys with nomenclature and designations.

Common name	UNS no.	Werkstoff nr.	ISO designation
Nickel 200	N02200	2.4060/66	NW2200
Alloy 214	N07214		
Alloy 330	N08330	1.4864	
Alloy 400	N04400	2.4360	NW4400
Alloy K-500	N05500	2.4375	NW5500
Alloy 600	N06600	2.4816	NW6600
Alloy 686	N06686		
Alloy 718	N07718	2.4668	NW7718
Alloy MA754	N07754		
Alloy 800	N08800	1.4876	NW8800
Alloy 800HT	N08811	1.4876	NW8811
Alloy 901	N09901	2.4662	NW9911
Alloy 925	N09925		
Alloy B-2	N10665		
Alloy C-22	N06022		
Alloy G-3	N06985	2.4619	NW6895
Alloy C-276	N10276	2.4819	NW0276
RENÉ 41	N07041		

The European Standards system of identification, sometimes called the Werkstoff numbering system, defines rules for designating steels and alloys by a series of numbers and letters to express principal characteristics and applications. This identification system is international and incorporated into the national standards of the participating CEN members, including Austria, Belgium, Denmark, Finland, France, Germany, Greece, Iceland, Ireland, Italy, Luxembourg, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, and the UK.

See also: Nickel-based Superalloys: Alloying

Bibliography

European Committee for Standardization 1998 *CEN Report CR 10260 Designation System for Steels—Additional Symbols*. European Committee for Standardization, Central Secretariat, Brussels

European Committee for Standardization 1992 *European Standard EN 10027-1 Designation System for Steels. Part 1: Steel Names, principal symbols*. European Committee for Standardization, Central Secretariat, Brussels

European Committee for Standardization 1992 *European Standard EN 10027-2 Designation System for Steels. Part 2: Steel Numbers, principal symbols*. European Committee for Standardization, Central Secretariat, Brussels

Society of Automotive Engineers 1996 *Metals and Alloys in the Unified Numbering System*, 7th edn. Society of Automotive Engineers, Warrendale, PA

J. H. Weber
Special Metals Corporation, Huntington, West Virginia, USA

Nickel-based Superalloys: Alloying

The nickel–chromium and nickel–iron–chromium-base superalloys are critical materials for high-stress, high-performance applications in the aerospace, power generation, and transportation industries. Superalloys are noted for their ability to retain their high strength and creep-rupture resistance up to a high fraction of their melting temperature, in addition to generally high strength. These alloys are based on a face-centered-cubic (f.c.c.) matrix which is tolerant to significant alloying for strengthening and for resistance to oxidation and other environmental attack. The development of superalloy compositions depends on an understanding of the effects of the various alloying elements on the alloy microstructural features, and the role these features play in imparting the desired high performance.

Chromium additions of ~10–25% to nickel provide an alloy base that is both oxidation- and creep-resistant. Chromium oxide scales provide environmental protection of the underlying base alloy at high temperatures by restricting the diffusion rate of oxygen and sulfur inward, and metallic elements from the alloy outward. A more formable and machinable material is provided by additions of iron to the nickel-base alloy. These nickel–chromium and nickel–iron–chromium-base alloys have relatively low creep strength. However, the properties of both of these f.c.c. matrices can be greatly improved by solid-solution strengthening, precipitation strengthening, and dispersion strengthening.

Strong solid-solution strengthening of the γ-f.c.c. matrix is produced by additions of the refractory elements molybdenum and tungsten. This effect is noted at elevated temperatures due to the slow diffusion of these elements through the matrix structure. Aluminum and chromium are also potent solid-solution strengtheners. Iron, titanium, vanadium, and cobalt also contribute, but to a lesser degree. Solid-solution strengthening is related to differences in atomic diameter. An additional contribution can result from differences in electron vacancy number. The latter effect is partly due to a lowering of stacking fault energy, which would lead to reduced creep by making cross slip more difficult in the γ matrix. Superalloys having good creep strength due to solid-solution strengthening, combined with excellent resistance to oxidation and corrosion, include INCONEL alloy 617 and HASTELLOY alloy X.

The predominant strengthening precipitates in nickel-base superalloys are gamma-prime, γ', and gamma-double prime, γ''. The γ' phase, typically identified as $Ni_3(Al,Ti)$, is an intermetallic phase between nickel and aluminum + titanium. Its crystal structure is an ordered f.c.c., $L1_2$-type that contributes antiphase boundary (APB) strengthening to the two phase γ–γ' alloy. Because the structure of both the γ' precipitate and the γ matrix are f.c.c. and the mismatch between the phases is low (0–1%), the interfacial energy is low and long-term stability of the precipitate is enhanced. The γ' phase has two unusual characteristics. First, its strength increases with increasing temperature to ~800 °C, and second, it has a limited amount of inherent ductility that prevents severe embrittlement should a massive morphology, such as a grain boundary film, develop. The γ' phase can be highly alloyed. The elements aluminum, titanium, molybdenum, tungsten, and chromium strengthen the γ' phase. Niobium and tantalum also strengthen it, but to a lesser extent.

The γ'' phase is also an ordered phase but has a body-centered-tetragonal structure and the composition Ni_3Nb. While γ' is the predominant matrix-strengthening phase in nickel-base superalloys, addition of significant levels of iron causes two changes. First, the γ'' phase tends to be more stable than γ' and hence the γ'' serves as the major strengthening precipitate, although γ' can still be present. Compositions representing this situation include

INCONEL alloys X750, 706 and 718. Second, the η phase in nickel–iron-base superalloys is primarily Ni_3Ti and aluminum plays essentially no direct role in the precipitation strengthening. However, the stability of both γ' and γ'' phases in Ni–Fe superalloys is enhanced by aluminum.

Long-time exposure at about 650–750 °C can cause transformation of the γ' and γ'' phases to η and δ phases, respectively. Coherency strain energy associated with the mismatches in lattice constants is believed to be the driving force for these transformations. The η phase has a h.c.p. crystal structure and forms either as cellular precipitates or intragranular Widmanstätten plates. Cellular precipitation that nucleates at the grain boundaries can be detrimental to stress rupture ductility. The second mode, Widmanstätten plates may reduce strength, but not ductility. The transformation of the metastable γ'' phase to the equilibrium δ, with a DO_2 crystal structure, as well as the η phase formation can be slowed by: (i) additions of aluminum and boron, (ii) replacement of some niobium with tantalum, and (iii) limiting the temperature of application.

Creep deformation occurs in crystalline solids by three mechanisms: dislocation creep, diffusional creep, and grain boundary sliding. The volume fraction of precipitate particles has a large effect on creep and stress rupture resistance in nickel-base superalloys. Creep strength, as defined as the stress producing a given steady-state creep rate, has been demonstrated to be proportional to the γ' volume fraction. In addition, stress rupture strength of many commercial superalloys with γ' contents of 15–60 vol.% have been shown to be proportional to the γ' volume fraction. The γ' volume fraction, particle size, and morphology affect the interaction between the precipitates and deformation-related dislocations. The γ' precipitates are cut by dislocations when the interparticle spacing is small, while at larger particle spacing dislocations bypass γ' by looping or climb mechanisms. The mode and extent of creep deformation and the transition from the cutting to bypassing also depend on deformation strain rate and temperature, APB and fault energy of the γ', the strength of the γ matrix and γ' phases, coherency strains, and solute diffusivity in γ and γ'. All of these factors are a function of the alloy composition. The γ' volume fraction can be raised by increasing the aluminum, titanium, tantalum, and niobium contents; it can also be increased by reducing the solubility of these elements in the γ matrix by raising the chromium, cobalt, molybdenum, and tungsten contents in the alloy.

Grain boundaries are relatively weak links in nickel-base superalloys. A reduction in grain boundary sliding obtained by small additions of elements such as carbon, boron, zirconium, and hafnium can, thus, have a significant effect on creep strength. Carbon additions result in precipitation of $M_{23}C_6$ (usually $Cr_{23}C_6$) particles, the distribution of which at grain boundaries can be controlled through heat treatments. A continuous grain boundary film of $M_{23}C_6$ can form during prolonged exposure above about 975 °C. The reaction to form this film occurs between intragranular MC carbides, rich in Ti, Hf, Zr, and refractory elements, and the γ matrix to form γ' plus $M_{23}C_6$ at grain boundaries. The presence of this film reduces rupture ductility. Boron, zirconium, and hafnium segregate to grain boundaries and improve high-temperature creep strength and ductility. These grain-boundary-segregating elements appear to reduce grain boundary diffusion rates. This would lower creep rates and slow down the process of intergranular stress rupture. Hafnium additions, up to about 2.0%, can also modify carbide morphology and improve rupture ductility. Some newer superalloys have a portion of the carbon replaced by boron and use borides, which, unlike primary MC carbides, can be manipulated by heat treatment. Higher boron and lower carbon contents are also used to manipulate the formation of the relatively ductile γ–γ' eutectic phase.

The nickel-base alloys with the highest creep resistance have the highest γ' volume fractions and high levels of refractory elements. These alloys, employed in high- temperature, high-stress turbine components such as blades, are used in the as-cast and heat-treated condition. Examples of such alloys are alloy B-1900, MAR M-200, and alloy IN-738. Large turbine components that cannot be cast to shape are made from wrought superalloys with lower γ' volume fractions and lower strength so that the alloys can be mechanically worked by rolling or forging. These alloys, such as Waspaloy, UDIMET alloy 700, and NIMONIC alloy 115, have higher chromium and lower refractory element contents to avoid formation of unwanted topologically close-packed phases that degrade properties. The creep strength of wrought alloys is lower than that of the cast alloys. Part of this difference is attributable to grain size; cast alloys have a coarser grain size than wrought alloys.

Manipulation of the grain structure of superalloys has produced significant improvement in creep strength. Directional solidification (DS) technology has minimized or eliminated grain boundaries oriented transversely to the applied stress and resulted in improved creep strength compared with polycrystalline cast alloys. Extension of this technology to produce single crystal alloys has improved creep strength further. The latter improvement has allowed the elimination of elements such as boron, zirconium, and hafnium that were previously added to improve the grain boundaries. Examples of these directionally solidified and single crystal alloys include DS MAR M-200 + Hf and alloy CMSX-4, respectively.

Powder metallurgy processing is also used for the production of superalloys. Powders produced by rapid solidification of molten alloy are subsequently

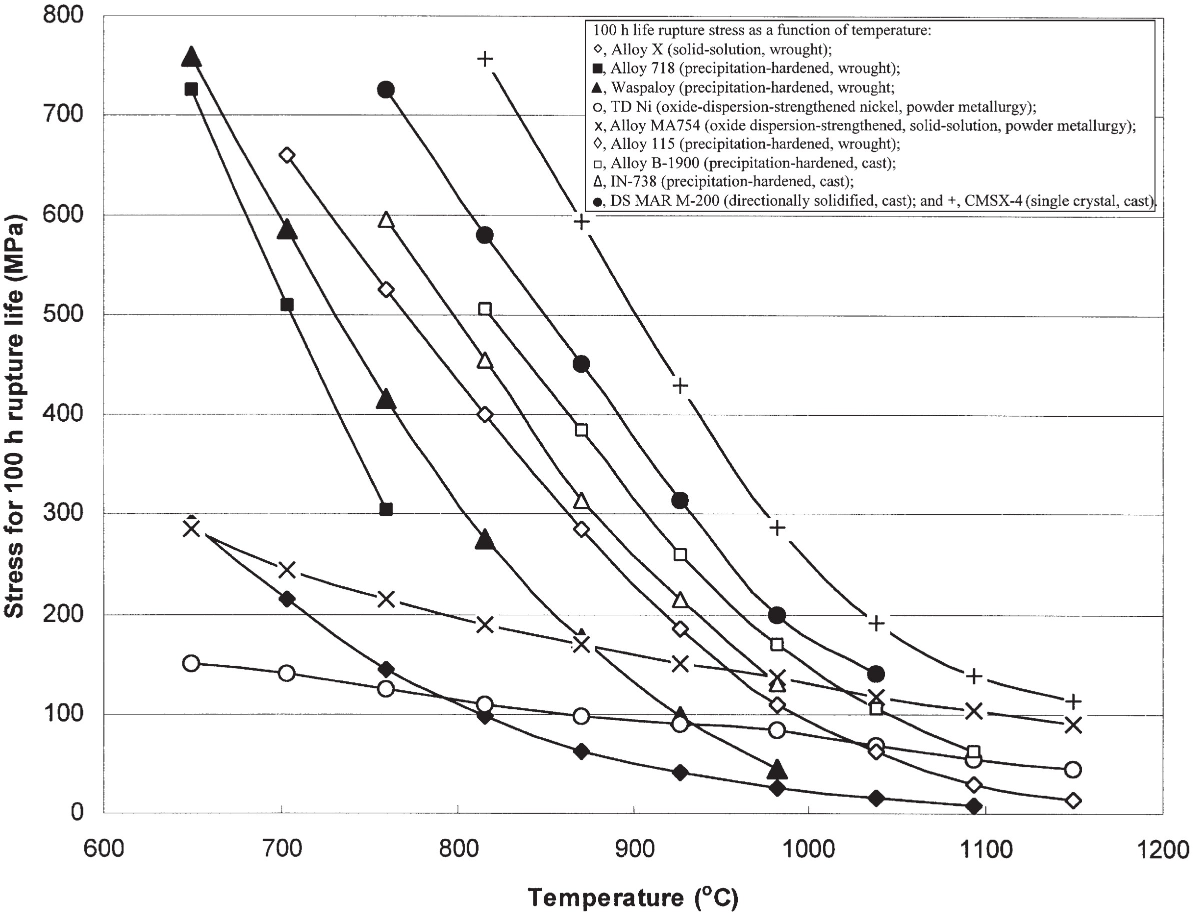


Figure 1
Creep strengths of selected superalloys.

consolidated to produce alloy structures that are fine-grained and amenable to hot-working practices. Very strong alloys with cast-type compositions can now be treated as wrought alloys. In addition, the maturation of hot isostatic pressing as a consolidation technique has allowed near-net shapes to be manufactured with the attendant benefit of cost reduction.

Powder metallurgy processing of a different type has produced superalloys with greatly improved creep strength by incorporating uniform dispersions of insoluble fine particles in superalloy matrices. Chemical techniques have been used for some time to produce alloys without reactive elements. Unfortunately, these reactive elements such as aluminum and titanium are also the ones providing precipitation hardening and good creep strength. Mechanical powder production techniques, for example, mechanical alloying, have overcome this limitation and allowed combinations of solid-solution strengthening, precipitation hardening, and dispersion strengthening to be achieved in a single alloy. Examples of oxide dispersion-strengthened (ODS) alloys are INCONEL alloys MA754 and TD NiCr. As with DS materials, anisotropic and aligned grain shapes are critical for maximizing creep strength in these alloys.

Advanced nickel-base materials more recently developed include composite superalloys and intermetallics. The principle of composites is to strengthen a ductile nickel-base matrix with strong second-phase fibers. This can be achieved either by mechanically incorporating fibers, such as tungsten, into a superalloy matrix or by directionally solidify liquid eutectic alloys such that one phase grows with a fibrous or lamellar morphology. Studies of intermetallics have been directed to the nickel aluminides NiAl and Ni_3Al and alloying of these bases to achieve useful combinations of strength and ductility. As expected, alloys based on the NiAl intermetallic also exhibit good oxidation and corrosion resistance.

The creep strengths of many of the types of superalloy discussed above are compared in Fig. 1. Stress-to-rupture in 100 h is plotted as a function of test temperature. The increases in stress rupture strength when going from solid-solution strengthened alloys to alloys strengthened by increasing fractions of γ', and finally to single crystal and ODS alloys are shown.

See also: Nickel Alloys: Nomenclature

Bibliography

Benjamin J S 1970 Dispersion strengthened superalloys by mechanical alloying. *Metall. Trans.* **1**, 2943–51

Decker R F 1979 Strengthening mechanisms in nickel-base superalloys. In: Bradley E F (ed.) *Source Book on Materials for Elevated-temperature Service*. American Society for Metals, Metals Park, OH, pp. 275–98

Desforges C D 1979 Metals and alloys for high temperature applications: current status and future prospects. In: Bradley E F (ed.) *1979 Source Book on Materials for Elevated-temperature Service*. American Society for Metals, Metals Park, OH, pp. 1–18

Gell M, Duhl D N 1986 The development of single crystal superalloy turbine blades. In: Allen S M, Pelloux R M, Widmer R (eds.) *Advanced High Temperature Alloys: Processing and Properties*. American Society for Metals, Metals Park, OH, pp. 41–9

Tundermann J H, Tien J K, Howson T E 1996 Nickel and nickel alloys. In: *Kirk–Othmer Encyclopedia of Chemical Technology*, 4th edn. Wiley, New York, NY, Vol. 17, pp. 1–17

Westbrook J H 1993 Structural intermetallics: their origins, status, and future. In: Kroschwitz S I, Howe-Grant M (eds.) *Structural Intermetallics*. The Minerals, Metals, and Materials Society, Warrendale, PA, pp. 1–15

J. H. Weber
Special Metals Corporation, Huntington,
West Virginia, USA

This Page Intentionally Left Blank

O

Oxide Glasses

The structure of the typical glass formers B_2O_3, SiO_2, GeO_2, and P_2O_5 is reported in terms of three-dimensional infinite cross-linking of units forming short-range order with no long-range order. The effects of intrinsic defect formation and of extrinsic defects caused by foreign impurities are discussed in relation to basic properties of the glassy state. Glasses for practical purposes result from the many possible combinations of network-forming oxides together with one or several modifier oxides M^I_2O ($M^I = $ Li, Na, or K) and/or $M^{II}O$ ($M^{II} = $ Mg, Ca, Sr, Ba, Pb) and/or $M^{III}_2O_3$ ($M^{III} = $ La, Ce). Incorporating the so-called intermediate oxides, Al_2O_3, Fe_2O_3, TiO_2, ZrO_2, etc., into glasses leads to an enormous extension of possible vitreous materials with special properties. Ion conductivity and optical properties are briefly reviewed in this article.

1. *Glass-forming Binary Oxides*

In the broad field of solid-state materials, the binary oxides B_2O_3, SiO_2, GeO_2, and P_2O_5 are known as typical glass formers. Using many types of methods of preparation, they form readily the noncrystalline state. With sufficient purity, an isotropic bulk body is easily formed even as the result of slow cooling from the melt or by the application of sol–gel preparation techniques, or as thin films by deposition from the vapor state.

1.1 The Glass Structure as a Random Network of Units

The increased glass-formation tendency of the oxides is a result of fundamental features of the chemical bond. Starting from the left-hand side of the periodic table, owing to the high difference of the electronegativity, highly connective crystalline ionic lattices of oxides prevail, yielding packing of units with a high density. Going on to the right-hand side, an increased contribution of covalent bonding occurs, and at the same time the coordination number decreases. Especially in the upper series, the bonds are more directed to the oxygen ligands. Hence, space filling is reduced because structures are formed which can be considered as intermediate with regard to the molecular structures of the volatile oxides of the nonmetallic elements at the upper right-hand corner of the periodic table. Of course, owing to the ability to form two-fold covalent bonds, oxygen is specially suited to link structural units having coordination numbers of 3, 4, or 6, and which makes possible continuous cross-linking, thus leading to highly polymeric network structures. The requirements of electroneutrality, allowing at most little charge separation, suggest the formation of triangular units of $BO_{3/2}$ or of tetrahedral units such as $[BO_{4/2}]^-$, $SiO_{4/2}$ and $GeO_{4/2}$, $OPO_{3/2}$, or $[PO_{4/2}]^+$. Zachariasen (1932), the founder of the theory of glass structure, defined these units to be the origin of a molecular-like short-range order. Owing to three-dimensional infinite cross-linking, highly polymeric network structures are formed without or with only very poor residual order in the adjacent coordination spheres.

However, units with coordination number of 6, e.g., the corner-shared octahedrons $ReO_{6/2}$ of ReO_3 or of compounds with the perovskite-type structure, do not seem to be able to form glasses. This appears to be the result of a strong directionality of the bonds with angles near to 180° at the bridging oxygens. Hence, the distribution of spatial configurations in the adjacent coordination spheres is extremely restricted, thus leading to the preference for crystallization.

Figure 1(a) shows for SiO_2 that the bond angle, α, at the bridging oxygen atom is $\alpha < 180°$. Variation of such angles around an average value of 150° and the freedom of rotation about the Si–O bonds make possible the occurrence of a broad distribution of structural configurations, yielding a continuous random network structure with no, or strongly reduced, long-range order. The essential structural prerequisite for a preference of glass formation is the existence of units with coordination numbers of 2, 3, or 4, which are connected by a divalent linking atom, e.g., oxygen. As shown in Fig. 1(a), the structure of vitreous silica can also be described by randomly distributed five-, six-, seven-, and higher-membered rings. Their vibrational dynamics have been studied by Elliott (1997).

Depending on the kind of preparation and thermal treatment, frequently a small residual degree of medium-range order is found to be maintained. In comparison to a random distribution, just those configurations in the second and sometimes even in the third coordination sphere appear with an increased frequency, which are typical for the crystalline modifications tridymite and cristobalite of silica. Actually, analysis of experimental radial distribution function results suggests small contributions of those configurations, which reflect a relationship to the crystallite approach to glass structure as proposed by Lebedev (1921). The structure of vitreous GeO_2 is analogous to that of vitreous silica, having an average angle at the bridging oxygen atoms of 133°.

Figure 1(b) shows the two-dimensional connection of rings of $BO_{3/2}$ units in vitreous B_2O_3 with random ring statistics and the preferential formation of boroxol rings, as deduced from diffraction measurements (Bell and Carnevale 1981).

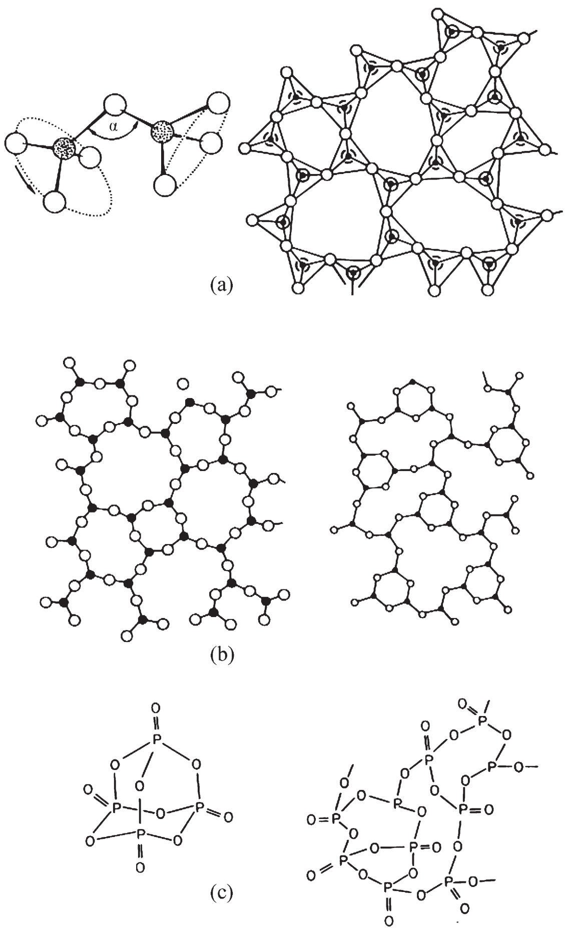


Figure 1
Schematic illustration of the network structure of the glass-forming oxides (a) SiO_2 and GeO_2, (b) B_2O_3, (c) P_2O_5.

Mixtures of B_2O_3 and SiO_2 have been shown to form glasses by using sol–gel preparation techniques. Commonly, small additions of a third, more basic, oxide initiate phase separation in the melt. Vogel (1992) elucidated the manifold phenomena of phase separation in glasses. The combination of B_2O_3 with P_2O_5 leads to the crystalline compound BPO_4, whose structure consists of a network of alternating $(BO_{4/2})^-$ and $(PO_{4/2})^+$ tetrahedrons. The structure and sequence of crystalline modifications occurring on heating BPO_4 is very analogous to that for silica; however, with the essential difference that glass formation is completely inhibited. Indeed, the energetically favored pairwise contacts of $(BO_{4/2})^-$ or $(PO_{4/2})^+$ units imply the formation of even-membered rings and, obviously, such a restriction is the reason for the lack of glass formation.

P_2O_5 in the pure state exists as the molecular compound P_4O_{10}, whose structure is closely related to the urotropine structure. Contact with protic agents, e.g., traces of water, gives rise to polymerization leading to two-dimensional continuous random cross-linking of $OPO_{3/2}$ units in the glassy state (Fig. 1(c)).

The fundamental structural features for glass formation are also valid for chalcogenide glasses. Because of their chain-like structure, elementary sulfur and selenium and also their binary compounds, e.g., with germanium and arsenic, act as glass formers. Many combinations have been studied (for a review see Feltz 1993).

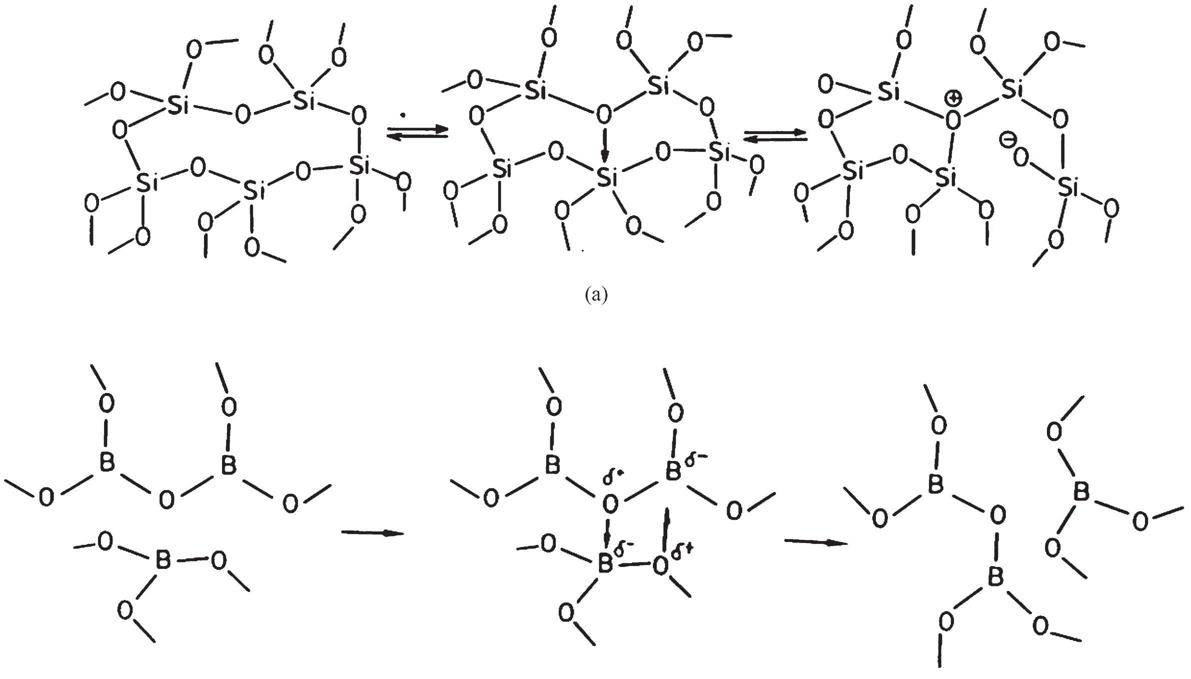
1.2 Intrinsic and Extrinsic Defects

The ideal structure of glass-forming oxides and chalcogenides is understood as a continuous random network of subassemblies in a structural skeleton that, although variable in its spatial configuration, becomes rigid when all the bonds are complete. With decreasing temperature, the melts of glass-forming compounds undergo a continuous transition from a viscous liquid to the vitreous state showing the typical shape-conserving behavior of an elastic solid. Of course, increasing the viscous flow of a highly polymeric structure requires the breaking of bonds. Partial depolymerization of the continuous network structure has to take place the higher the temperature, thus leading to intrinsic point defect formation. Actually, the nonbonding electron pairs of the bridging oxygen atoms make possible a favored mechanism of defect formation. As shown in Fig. 2(a), the covalent bonds are preferentially broken in such a way that the electron pair of the corresponding bond remains on one of the fragments; in other words, this is a heterolytic process. The energy required for breaking the bond is reduced by an increase of the coordination number at a silicon atom in the transition state. Effectively, the vacant site is filled by a nonbonding electron pair from another neighboring bridging oxygen.

Such an increase of the coordination number from 4 to 5 at the silicon atom in the transition state of the defect formation reaction is quite plausible (Liebau 1985). The lone pair electrons of a bridging oxygen from the environment guarantees a stabilization reducing the free energy of formation of pairs of charged dangling bonds as intrinsic defects. The positively and negatively charged defect centers should thus be assumed to occur in the supercooled melt with an equilibrium concentration $n_C = n_0 \exp(-\Delta G/2RT)$ at a given temperature. Its value is then determined by the free energy of the relevant defect generation reaction

$$2SiO_{4/2} + O(SiO_{3/2})_2 \rightleftharpoons {}^{+}O(SiO_{3/2})_3 + SiO_{3/2}O^-$$

The total concentration of bonds is given by $n_0 = (\rho N_A/\bar{M})\Sigma x_i z_i/2$, where $\bar{M}$ is the average atomic mass, N_A is the Avogadro number, x_i is the mole



(a)

(b)

Figure 2
Intrinsic constitutional disorder of vitreous silica with increase of coordination number at silicon in (a) the transition state and (b) the interlayer interaction of vitreous B_2O_3.

fraction of the corresponding element, z_i the coordination number, and ρ is the density.

A concentration, n_C, of defect pairs corresponding to the temperature of the glass transition range is frozen in during the process of vitreous solidification. Such centers with a concentration of about $10^{19}\,cm^{-3}$ persist in the vitreous state of silica (Lucovsky 1979) because it is quite unlikely that all the broken bonds can be reunited into intact units to form an ideal glass structure without defects during solidification of the supercooled melt. Consequently, homogeneous glasses of stoichiometric composition exhibit a constitutional disorder owing to the presence of positively and negatively charged defect centers. In analogy to point defects in crystals, the predominant defect type is that whose formation reaction involves the smallest expense of free energy.

Foreign impurities, e.g., small concentrations of water or Na_2O, are expected to reduce the concentration of intrinsic defects corresponding to the reactions

$$^{+}O(SiO_{3/2})_3 + SiO_{3/2}O^- + H_2O \rightarrow$$
$$2SiO_{3/2}OH + O(SiO_{3/2})_2$$

$$^{+}O(SiO_{3/2})_3 + SiO_{3/2}O^- + Na_2O \rightarrow$$
$$2SiO_{3/2}O^- + 2Na^+ + O(SiO_{3/2})_2$$

SiO_2 prepared by $SiCl_4/O_2$ pyrolysis contains residual Si–Cl bonds. Increasing oxygen cleavage occurs when silica is heated to temperatures above about 1200 °C. SiO_2 is decomposed on volatilization: $SiO_2 \rightarrow SiO + \frac{1}{2}O_2$. Therefore, deposition from the vapor at a reduced oxygen pressure leads to nonstoichiometric $SiO_{2-\delta}$. Oxygen defect states $\uparrow SiO_{3/2}$ with single electrons prevail. They can be interpreted as homolytically broken Si–Si bonds. At higher oxygen defect concentrations, a range of $Si(Si_yO_{4-y})$ $(0<y<4)$ species is formed, leading to increasing phase separation on annealing, which is in accordance with the equilibrium phase diagram of Si/SiO_2.

The rutile-type analogous GeO_2 modification with octahedral $GeO_{6/3}$ units is transformed at 1049 °C to a β-quartz-like structure with $GeO_{4/2}$ units, yielding already at 1116 °C a highly viscous melt. The glass that is formed on cooling has a similar structure to that of silica. There is also a concentration of pairs of oppositely charged defects as shown in Fig. 2(a) for silica; however, with the difference that the transition state with increased coordination is expected to be more stabilized. In contrast to vitreous silica, an increased coordination number of 5 or even 6 at the germanium atom should be favored because for germanium the empty *d* electron states are energetically more available. Further, oxygen liberation is facilitated when

GeO_2 melts are heated in a reduced oxygen atmosphere. The atomization enthalpy of GeO_2, $\Delta H_A = 1415\,\mathrm{kJmol^{-1}}$, is significantly lower than that for SiO_2, $\Delta H_A = 1855\,\mathrm{kJmol^{-1}}$. Up to an overall composition of $GeO_{1.8}$, melting of Ge/GeO_2 mixtures in an argon atmosphere leads to homogeneous glasses even at slow cooling. The formation of charged defect pairs $[GeO_{3/2}]^-$ $[O(Ge(O_{3/2})_3]^+$ is suggested to be energetically more preferred than neutral defects of $GeO_{3/2}$ or units of $O_{3/2}Ge–GeO_{3/2}$. The latter type of structural units is found to comprise the characteristic short-range order of the chalcogenide glasses Ge_2S_3 and Ge_2Se_3 (Feltz 1977).

Because of the well-known Lewis acidity of trivalent boron, vitreous B_2O_3 is expected to show an increased stability of the transition state, which is shown schematically in Fig. 2(b). When a bridging oxygen atom with its nucleophilic lone pair electrons is close to a boron atom of an adjacent unit, the trigonal planar $BO_{3/2}$ group will form a distorted tetrahedron, thus strengthening the structure by an interlayer interaction. However, commonly because of hygroscopicity, the structure of B_2O_3 glasses is modified by OH groups forming $HOBO_{2/2}$ units, which reduces the interlayer interaction or even interrupts the two-dimensional network structure:

$$O_{3/2}B - O(BO_{2/2})_2 + H_2O \rightarrow BO_{3/2} + 2HOB(O_{2/2})_2$$

P_2O_5 is one of the most hygroscopic materials. Hence, the real structure of vitreous phosphorus oxide is governed by the formation of OH groups which interrupt the network of two-dimensional cross-linking of the $O{=}PO_{3/2}$ units. Actually, the structure can be interpreted as a highly condensed poly-phosphoric acid with randomly distributed cross-linking between the chains. At high purity, partial formation of charged intrinsic defects $[PO_{4/2}]^+$ and $O{=}PO_{2/2}O^-$ in the polymeric state should also be possible.

1.3 Viscous Flow and the Glass Transition Temperature

Only the presence of extrinsic and/or intrinsic defects makes possible the viscous flow of typical glass formers such as the oxides. Of course, the concentration of intrinsic defects and those resulting from oxygen cleavage increase with temperature. The concentration of residual OH groups can be reduced by the evaporation of H_2O, thus leading to an increase of the connectivity because bridging oxygens are restored. Actually, there are several mechanisms of defect formation with consecutively increasing values of the free energy, thus leading to increasing values of the activation energy of viscous flow with increasing temperature. This is shown in the Arrhenius plot of viscosity $\eta = \eta_0 \exp(E_v/RT)$ in Fig. 3.

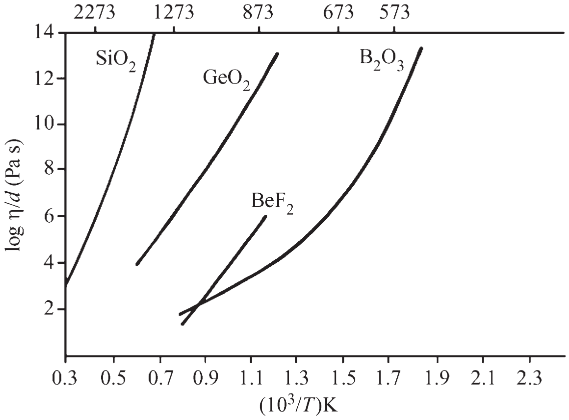


Figure 3
Temperature dependence of the viscosity of various glass-forming compounds.

Depending on time, there can be a continuous transformation from the glassy state showing the typical elasticity of solids to the viscous flow of a liquid. Hooke's law states that a change of shear stress, $\mathrm{d}S_G$, will be proportional to the increase of shear deformation, $\mathrm{d}\alpha$, which gives the time dependence

$$\frac{\mathrm{d}\alpha}{\mathrm{d}t} = \frac{1}{G}\frac{\mathrm{d}S_G}{\mathrm{d}t}$$

where G is the shear modulus. For plastic deformation of the vitreous body, i.e., with the onset of viscous flow, the shear stress, S_G, will be proportional to the gradient of the velocity of flow perpendicular to the direction of shear stress. This corresponds to a Newtonian liquid satisfying the equation

$$\frac{\mathrm{d}v}{\mathrm{d}x} = \frac{S_G}{\eta}$$

where η is the viscosity measured in Pas. The total deformation $(\mathrm{d}\theta/\mathrm{d}t)$ is given by Maxwell's differential equation for viscoelastic flow behavior:

$$\frac{\mathrm{d}\theta}{\mathrm{d}t} = \frac{1}{G}\frac{\mathrm{d}S_G}{\mathrm{d}t} + \frac{S_G}{\eta} \quad \text{or} \quad \frac{\mathrm{d}S_G}{\mathrm{d}t} = G\frac{\mathrm{d}\theta}{\mathrm{d}t} - \frac{S_G}{\eta/G}$$

The rate of variation of the shear stress $(\mathrm{d}S_G/\mathrm{d}t)$ is proportional to the rate of deformation $(\mathrm{d}\theta/\mathrm{d}t)$. If flow processes are operative, the rate of shear stress coupled with deformation will become smaller the greater is S_G and the smaller is the quotient η/G. The latter represents a relaxation time $\tau = \eta/G$ describing the rate at which the body escapes the external load by decay of the shear stress. Thus, according to this equation, whether the material in question is to be regarded as a glass or a supercooled liquid depends on the duration, Δt, of the observations. The term S_G/τ can be neglected if $\Delta t \ll \tau$, because the body is actually

Hookean. In the opposite case, i.e., $\Delta t \gg \tau$, glasses and even crystalline solids can undergo plastic deformation, i.e., $d\theta/dt \approx 0$. In the intermediate region, i.e., for $\Delta t \approx \tau$, the transition from the glassy solid to the liquid state or vice versa, the so-called glass transition, can be observed. With the arbitrarily taken relaxation time of about 5 min as a convenient time of observation, the shear modulus of 31 GPa of vitreous silica provides a viscosity $\eta = 10^{13}$ Pas, which is a characteristic value of the viscosity in the glass transition range. The glass transition temperature of about 1150 °C for vitreous SiO_2, 845 °C for glassy GeO_2, and in particular the low value of about 220 °C for B_2O_3, all depend strongly on the residual concentration of extrinsic defects.

Figure 4 shows the enthalpy as a function of temperature for glass-forming systems. If a glass melt is cooled with a rate, q_A, sufficient to prevent crystallization, a thermal equilibrium still exists in the region of supercooling. Despite the high viscosity, initially the redistribution of internal energy or enthalpy between the degrees of freedom of the system is able to keep pace with the cooling rate, $q_A = -(dT/dt)$. The number and type of ruptured bonds and thus the size of flow units, as well as their translational, vibrational, and rotational behavior, are subject to change with temperature, thus indicating structural relaxation. Delays are observed if the reaction rate, which is limited by such processes, falls behind the cooling rate. The system then can no longer keep pace with the cooling rate. The amount of enthalpy remaining in the melt is greater than that expected for thermodynamic equilibrium at that temperature. Consequently, the supercooled melt becomes additionally metastable because the structural relaxation process freezes in the transition range yielding the vitreous state. Glasses are metastable with respect to the supercooled melt and also with respect to the crystalline state, and are thus metastable in a dual sense. Glasses are said to be in the frozen state of a supercooled melt. The freezing temperature, T_{ZA}, is defined as the point of intersection of the two straight lines shown in Fig. 4 and, obviously, it depends on the cooling rate. Hence, rapid cooling with $q_B > q_A$ provides a higher freezing temperature $T_{ZB} > T_{ZA}$.

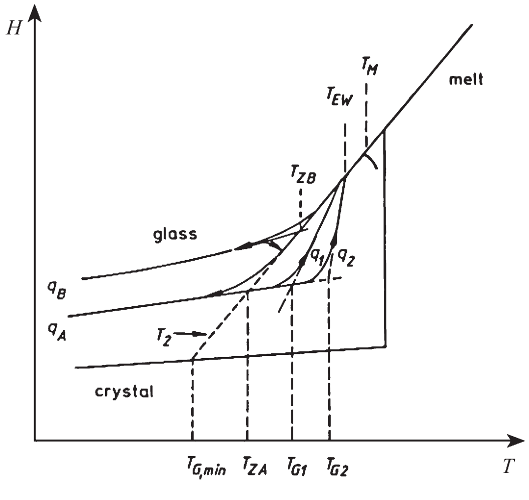


Figure 4
Temperature dependence of the enthalpy in the glass transition region.

Commonly, the glass transition is measured by thermal analysis or using a dilatometer starting from the glassy state with a heating rate $q_B > q_1$, $q_2 > q_A$. Because of retarded relaxation, the freezing temperature, T_{ZA}, will now be bypassed until the slope of the curve increases, the greater the slope, the higher the heating rate $q_2 > q_1$. The glass transition temperature T_{G1} or T_{G2} can be obtained from the tangents to the curves. Of course, its value is a function of the cooling rate as well as of the heating rate. For conventional practical purposes, the glass transition temperature corresponds to values of the viscosity in the range $10^{11} < \eta < 10^{13}$ Pas. It is only at the temperature T_E, the softening temperature being coupled with a value of the viscosity $\eta = 10^{10.3}$ Pas, that the behavior coincides with the portion of the curve that characterizes the state of equilibrium of the supercooled melt. T_M is the point of onset of change in shape owing to the weight of the specimen. Measuring a rapidly cooled sample, e.g., with $q_1 < q_B$, leads to a negative deviation from the curve before the freezing temperature T_{ZB} is achieved because there is sufficient time for structural relaxation. The whole thermal history of a solid in the vitreous state affects the behavior approaching the glass transition range and that is valid for all properties of glasses. Hence, the properties, e.g., the refraction index, can be tuned by thermal treatment within small tolerances, which is important for the application of glasses in optics.

2. Silicates, Germanates, Phosphates, and Borates as Glass-forming Systems

The great range of glass-forming systems results from the many possible combinations of different glass formers, and the variability of composition that is introduced by including some of the numerous other oxides.

There are substances with defined composition and structure, which undergo decomposition during melting, yielding a mixture of different compounds. Madrell's salt, crystalline $NaPO_3$, for example, has an infinitely extended helix of tetrahedral $OPO_{2/2}O^-$ units, with a repeat period of three units. In the melt, the structure undergoes considerable rearrangement. Cyclophosphates $(NaPO_3)_n$ with $3 < n < 7$ have been detected following vitreous solidification. Pb_2SiO_4 has the structure of a tetracyclo-silicate, according to the formula $[Pb_2O]_4[Si_4O_{12}]$ which is decomposed during melting, yielding a mixture of different cyclic and chain-like anion species, whose negative charge is compensated for by fragments of the modifications of PbO. The thermal cracking, yielding a mixture of different structural configurations, makes it possible to understand the formation of highly refractive flint glasses for optical purposes containing up to 70 mol% PbO.

The addition of the more basic metal oxides M^I_2O ($M^I = $ Na or K) and/or $M^{II}O$ ($M^{II} = $ Ca, Sr, Ba, Pb) to the usual glass-forming oxides B_2O_3, SiO_2, GeO_2, and P_2O_5 modifies the network structure. As shown schematically in Fig. 5, the modifier oxides Na_2O and CaO break the bridging bonds by nucleophilic attack, simultaneously generating negatively charged terminal (nonbridging) oxygen atoms, which preferentially coordinate the inserted metal ions. At the beginning, these interactions primarily have the character of ionic bonds, which means that the network is interrupted and the degree of connectivity is progressively degraded by such additives.

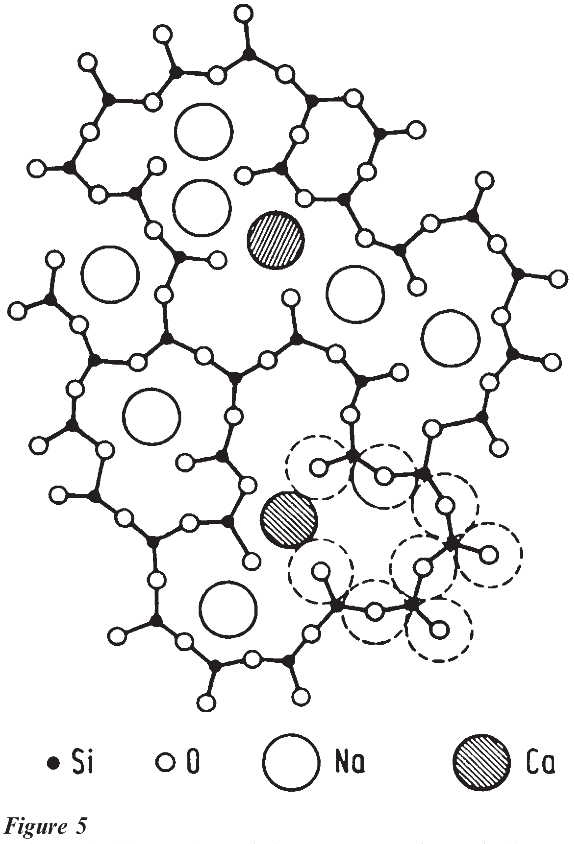


Figure 5
Schematic illustration of the structure of a soda-lime silicate glass (the fourth oxygens of the units directed upward or downward are omitted).

The glass-forming tendency usually increases with the incorporation of network modifiers because of the increase in structural configuration possibilities. Often it has a maximum in the region of the eutectic composition, which can be interpreted as a compositional region of inhibited crystallization owing to the disorder. However, sometimes even a special unit of short-range order is formed in the melt, which exists only in the noncrystalline state (Feltz and Pfaff 1985). Relatively high values of the viscosity are typical because of the lowered liquidus temperature in a eutectic region. However, even small amounts of additives can sometimes, at least locally, greatly reduce the high viscosity of network formers, e.g., traces of Na_2O at the surface of vitreous silica, so that crystallization processes are initiated.

The incorporation of network modifiers alters fundamentally the glass properties. Owing to the formation of less directed ionic bonds, the structural skeleton collapses into a closer packing. Hence, the molar volume will decrease and the glass transition temperature is lowered owing to the reduced degree of cross-linking. An increase of the polarizability arises from the negatively charged nonbridging atoms, and thus the anharmonicity of thermal vibrations also increases. Consequently, the coefficient of thermal expansion will increase.

The higher the content of network modifier in the glass, the higher is the concentration of nonbridging atoms. The increasing saturation of the electrophilicity of the units of the network-forming oxide leaves the nonbonding electron pairs increasingly liable to polarization by the cations. Hence, the covalent character of the metal–oxygen bonds increases, so that the corresponding coordination polyhedrons increasingly take on the role, together with the rest of the anions, of determining the structure and properties of the glass. This gives a plausible explanation of the existence of glasses with high metal oxide content, so-called invert glasses (Trap and Stevels 1959). With only 40 mol% SiO_2 in a glass $(Na_2O)_{15}(K_2O)_{15}(CaO)_{15}(BaO)_{15}$-$(SiO_2)_{40}$, a network of silicate units no longer exists. The low degree of covalency in the M–O bonds and the resulting relatively high mobility of the melts lead to a strong tendency towards crystallization. This tendency is opposed by the presence of several different metal oxides with differing coordination numbers and slightly varying bond distances, with an increase in configuration entropy.

Obviously, with a variation of the ratio of network former and modifier the kind of interaction is progressively changing, thus making understandable the ambiguous function of the less basic so-called intermediate oxides in the structure of glasses. If introduced in low concentration, Li_2O and MgO have a modifier function with a coordination number of 6 or more. At high concentrations and especially in the presence of more basic modifier oxides, e.g., of Na_2O or K_2O, they are able to enter the network, forming

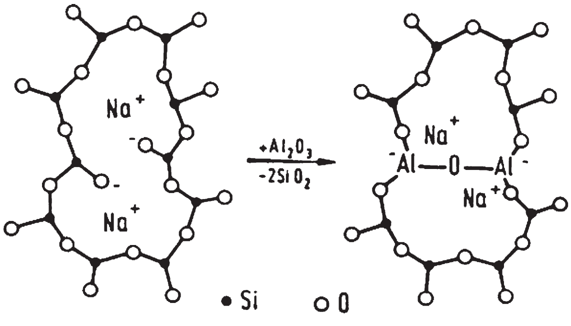


Figure 6
Schematic illustration of the replacement of two SiO_2 by Al_2O_3 in a sodium silicate glass (the fourth oxygens of the units directed upward or downward are omitted).

units of $(LiO_{4/2})^{3-}$ and $(MgO_{4/2})^{2-}$ with a coordination number of 4, i.e., strengthening of the silicate glass network occurs. In an analogous way, Al_2O_3 acts predominantly in strengthening. Al_2O_3 entering the structure of glassy SiO_2, GeO_2, or B_2O_3 with a coordination number 6, i.e. as a modifier, is found to be confined to small concentrations. As illustrated in Fig. 6, additions of alumina are able to compensate for changes of the glass structure and properties caused by the typical modifier oxides owing to the reaction

$$2SiO_{3/2}O^- + 2Na^+ + Al_2O_3 \rightarrow 2SiO_{4/2} + 2[AlO_{4/2}]^- + 2Na^+$$

Besides the $SiO_{4/2}$ units, tetrahedral units $[AlO_{4/2}]^-$ are formed, completing the network in a similar way to that known from the skeleton structure of feldspar. Fe_2O_3 shows a similar behavior. Other oxides of the intermediate type, e.g. TiO_2, ZrO_2, or Nb_2O_5, also form octahedral units that tend to strengthen the glass structure by the formation of strong bonds to the nonbridging negatively charged oxygens.

2.1 Silicate Glasses and Some Peculiarities of Germanate Glasses

Because of hygroscopicity, alkali silicate glasses are less important for practical purposes. In the series of binary systems M_2O–SiO_2 (M = Li, Na, K, Rb, Cs), the sodium silicate glasses show the highest value of ion conductivity. However, with gradual substitution of one alkali ion by another alkali ion, all properties that are affected by transport mechanisms display a distinctly nonlinear behavior. The conductivity of glasses containing a mixture, e.g., of sodium and potassium ions, is found to be lower by several orders of magnitude than that for the two pure-component systems. This effect is called the mixed-alkali or mixed-mobile-ion effect, indicating a mechanism of dynamic interaction between the mobile modifier ions and the network structure. Bond percolation, coupled with ion motion, has been suggested to be the reason for this unusual behavior (Ingram *et al.* 1991).

Extended ranges of phase separation in the liquid state are typical for the binary systems MO–SiO_2 (M = Mg, Ca, Sr) and also in the subliquidus range of temperature for BaO–SiO_2. More suitable for application are the combined systems M_2O–MO–SiO_2, e.g., Na_2O–CaO–SiO_2, used for windows or in optics as crown glasses. Figure 5 shows schematically the action of the two different oxides with a modifier function. Owing to the increased coulomb attraction of the divalent Ca^{2+} ions, the mobility of the Na^+ ions is generally blocked, i.e., the electrical conductivity is greatly reduced. However, they still tend to microphase separation, thus leading to leaching. An addition of 2–3% of alumina makes it possible to suppress this microphase separation, yielding glasses of high chemical stability that are suitable for many applications.

Despite the structure of vitreous GeO_2 being very similar to that of SiO_2, the interaction with modifier oxides gives rise to an anomaly in comparison with silicate glasses. In accordance with expectation, the structural modification of GeO_2, e.g., by Na_2O, is initially associated with an increase in density. For larger amounts of modifier the density passes through a maximum that is caused by a change of the germanium coordination number from 4 to 6. Obviously, the higher concentration of negatively charged nonbridging oxygens favors the formation of octahedral $(GeO_{6/2})^{2-}$ units that are incorporated into the network. Owing to the larger Ge–O distances, coupled with the higher coordination number, the density decreases until again a collapse of the network prevails caused by still higher concentrations of Na_2O. Of course, germanium more easily adopts a coordination number of 6 than silicon, which can also be seen considering the structure of $BaGe(Ge_3O_9)$ in comparison with that of benitoite $BaTi(Si_3O_9)$.

2.2 Borate and Borosilicate Glasses

Because of the high electrophilicity of the $BO_{3/2}$ units, owing to the three-fold coordinated boron, the glasses show an increased tendency for a change of the coordination number, giving rise to the so-called boron anomaly. Nonbridging oxygen formation as a result of Na_2O incorporation into B_2O_3 is found to be confined to less than 3 mol% modifier:

$$2BO_{3/2} + Na_2O \rightarrow 2BO_{2/2}O^- + 2Na^+$$

In the range up to about 40 mol% Na_2O, the formation of $[BO_{4/2}]^-$ units predominates. NMR measurements elucidated the continuous increase in concentration of these groups, yielding a maximum at 45 mol% (Bray 1985). They have a network cross-linking function because the connectivity increases.

Higher Na_2O contents produce a decrease in coordination number, until at 70 mol% all the boron is again in three-fold coordination. Obviously, owing to Coulomb repulsion, adjacent $(BO_{4/2})^-$ groups are less favourable than neighboring $BO_{2/2}O^-$ units, whose negative charge appears more localized on the nonbridging oxygens.

B_2O_3 is a common ingredient of chemically resistant and thermal shock-resistant glasses. Commonly, borosilicate glasses contain the components SiO_2 and Na_2O, i.e., nonbridging oxygens are thus present. By analogy with Al_2O_3, the incorporation of B_2O_3 into such a glass leads to the binding of the nonbridging oxygens providing tetrahedral units which are incorporated into the network, strengthening the glass structure.

2.3 Phosphate Glasses

Glass formation in the binary systems $M^I_2O–P_2O_5$ (M = Li, Na, K) and $M^{II}O–P_2O_5$ (M = Mg, Ca, Sr, Ba, Zn, Pb) from the melt exceeds a little the limiting composition MPO_3 or $M^{II}(PO_3)_2$ (50 mol% $M^{II}O$ or M^I_2O). At lower modifier concentrations, they are extremely reactive with water. Even $Ca(PO_3)_2$ glass still indicates a strong dependence of the optical absorption and dispersion behavior on a residual content of water between $0.005\,mol\,l^{-1}$ and $0.16\,mol\,l^{-1}$ (Ariagada *et al.* 1987). It is necessary to degrade the polymeric phosphate anion structure by sufficient addition of basic oxides in order to reduce the hydrolytic attack of water. As an example, the Na^+ superionic conducting system $Na_{1+3x}Zr_2(PO_4)_{3-3x}\text{-}(SiO_4)_{3x}$ with $0<x<1$ (NASICON) can be mentioned (Goodenough *et al.* 1976). The network structure is formed by octahedral $[ZrO_{6/2}]^{2-}$ units that are connected by tetrahedral $[P_{1-x}Si_xO_4]^{(1-x)+}$ units in the ratio 2 to 3. Hence, for $x=0$ the structure comprises a framework of $\{[(ZrO_{6/2})^{2-}]_2[(PO_{4/2})^+]_3\}^-$ with Na^+ inserted into the interstices of the network. Glasses having high Na^+ ion conductivity have been proposed based on this system (Susman *et al.* 1983).

3. *Ion Conductivity and Optical Properties of Oxide Glasses*

3.1 Ionic Conduction and the Structure of Glasses

Dielectric oxide glasses are frequently used as insulators, i.e., extremely high resistivity values and low dielectric losses are desirable. Figure 7 shows the high resistivity of different oxide glasses. The residual conductivity is limited by the mobility of Na^+ ions. Indeed, as a result of blocking owing to the presence of Ca^{2+} ions, glasses with higher Na^+ ion concentrations have a higher resistivity than vitreous silica containing only 10 ppmNa^+ (Mackenzie 1965).

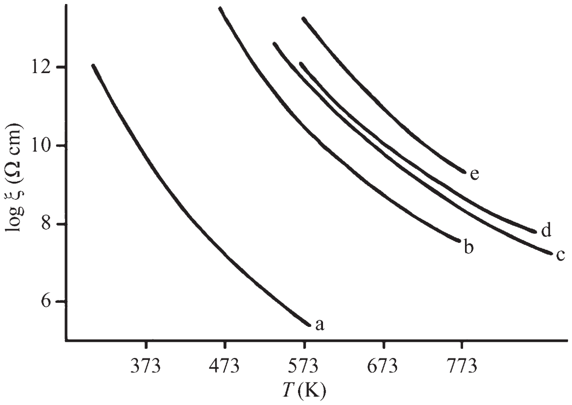


Figure 7
Resistivity plot vs. inverse temperature of highly resistive oxide glasses: (a) $Na_2O–CaO–SiO_2$ glass, (b) SiO_2 glass with 10 ppm Na^+, (c) $Ca_2Si_3O_8$ glass with 200 ppm Na^+, (d) $(CaO)_{0.43}(Al_2O_3)_{0.13}(SiO_2)_{0.42}$ glass with 10^3 ppm Na^+, (e) CaB_4O_7 glass with 5% Na^+.

Highly ionic conducting glasses have attracted increasing interest as solid electrolytes for batteries (Tuller and Barsoum 1985, Tuller *et al.* 1980, Feltz 1987, Burckhardt and Rudolph 1991). Only monovalent cations of a suitable size, e.g., Li^+ and Na^+ ions, have sufficient mobility and, of course, the conductivity is the higher the higher their concentration.

The mechanism has been described by ion jumping which involves a thermal activation of the mobility, i.e., all of the monovalent ions are assumed to be available for taking part in the transport and, owing to the loosened structure, the number of vacant sites in their surrounding is assumed to be sufficient. However, for $Na_2Si_3O_7$ in the glassy state, the Na^+ ion concentration is extremely high, e.g., $21\,mol\,l^{-1}$. Nevertheless, there are experimental results that suggest an alternative weak electrolyte model of ion jumping to be valid, even for low ion concentrations (Tomozawa *et al.* 1980, Lezikar *et al.* 1980). A temperature-dependent dissociation equilibrium is taken into account corresponding to

$$-ONa \rightleftharpoons -O^- + Na^+$$

for silicate, germanate or phosphate glasses, or

$$[(BO_{4/2})Na] \rightleftharpoons (BO_{4/2})^- + Na^+$$

for borate and borosilicate glasses, or

$$[(AlO_{4/2})Na] \rightleftharpoons (AlO_{4/2})^- + Na^+$$

for aluminosilicate glasses. Only the dissociated part of the total sodium ion concentration is assumed to take part in the conduction, i.e., the thermal activation energy is interpreted to be split into one term involving the thermal energy of dissociation and a second term

Table 1
Glasses with high Na^+ ion mobility.

Composition	$\sigma_{300\,^\circ C}$ ($S\,cm^{-1}$)	E_A (eV)	Stability against water	Reference
$Na_2Si_2O_5$	3.0×10^{-4}	0.63	Medium leaching	Feltz and Büchner (1987)
$(Na_2O)_{40}(B_2O_3)_{10}(SiO_2)_{50}$	2.0×10^{-3}	0.55	Medium leaching	Ingram and Hunter (1984)
$(Na_2O)_{40}(B_2O_3)_{50}(P_2O_2)_{10}$	1.8×10^{-4}		Medium leaching	Levasseur *et al.* (1981)
$(Na_2O)_{16.7}(Al_2O_3)_{16.7}(SiO_2)_{66.7}$	1.3×10^{-4}	0.94	Very stable	Hunold and Brückner (1980)
$(Na_2O)_{40}(Al_2O_3)_{15}(B_2O_3)_{15}(SiO_2)_{30}$	2.5×10^{-4}	0.64	Stable	Feltz and Popp (unpublished data)
$(Na_2O)_{33.3}(ZrO_2)_{16.7}(SiO_2)_{50}$	1.2×10^{-3}	0.55	Very stable	Susmann *et al.* (1983)
$Na_4Nb(PO_4)_3$	0.7×10^{-3}	0.60	Very stable	Barth and Feltz (1989)

that is the mobility activation energy of that part of the Na^+ ions that is liberated for motion.

The formation of negatively charged $(BO_{4/2})^-$ and $(AlO_{4/2})^-$ units as part of the network is coupled with the creation of more free volume for opening additional percolation paths. Because of the less localized negative charge, dissociative Na^+ excitation should be favored, i.e., the dissociation energy will be lowered. At the same time, the potential barrier profile along the percolation paths will become more balanced, producing lower values for the activation energy of the mobility. Altogether the Na^+ ion conductivity is suggested to be improved when negatively charged nonbridging oxygens are incorporated into network-forming units, as shown for borosilicate and aluminosilicate glasses, or by the action of other suitable intermediate oxides, such as ZrO_2 or Nb_2O_5. Some results are collected in Table 1. Impedance measurements have been used to reveal the mechanism of ion transport in glasses (Elliott 1988, Kahnt and Schirrmeister 1987, Kahnt 1991).

Ion transport is also used for the controlled formation of inhomogeneous glasses with a definite refractive index profile which are becoming increasingly important as components of gradient optical systems for imaging as well as for waveguide functions in micro-optics. Figure 8 shows the relative changes of the refractive index as a result of Na^+/Ag^+ interdiffusion (Kaps and Völksch 1987).

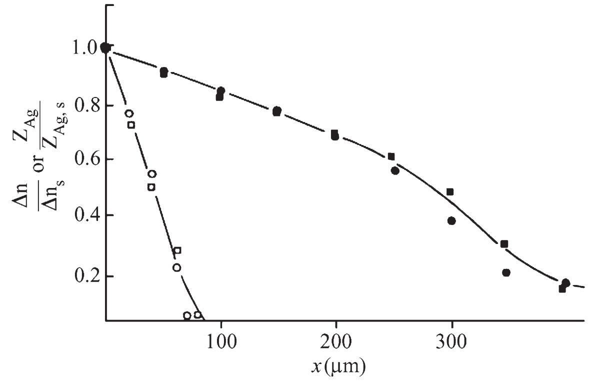


Figure 8
Relative changes of the refractive index $\Delta n/\Delta n_0 = (n_{x,t} - n_{x,0})/(n_{0,t} - n_{0,0})$ (○●) and relative Ag^+ concentration $z_{Ag}/z_{Ag,0}$ (□■), expressed by the counts in the microprobe detector after Na^+/Ag^+ interdiffusion at 283 °C (○□) and 423 °C (●■) during time of $t = 40$ h.

3.2 *Optical Properties of Oxide Glasses*

The refractive index and its dispersive wavelength dependence are important parameters that determine the suitability of oxide glasses as optical media. This theme is explored more thoroughly by Fanderlik (1983). For practical purposes, one can define a mean dispersion, originally introduced by Abbe, which is the difference in refractive indices for two wavelengths, $\Delta n = n_1 - n_2$. In the visible spectral region, it is also referred to as the main dispersion, defined as $n_F - n_C$, in terms of the refractive indices for certain spectral lines of cadmium, such as $F' = 479.99$ nm for n_1 and $C' = 643.85$ nm for n_2. The refractive behavior of an optical medium depends essentially on the reduced refractive index $n - 1$ in the relevant wavelength range, so that normalization of the refractive index n_e at a wavelength between F′ and C′ is necessary. The green line e = 546.07 nm of the mercury spectrum is specified for this purpose. The quantity $(n_{F'} - n_{C'})/(n_e - 1)$ is the relative dispersion, and the reciprocal of the relative dispersion is known as the Abbe number, $v_e = (n_e - 1)/(n_{F'} - n_{C'})$. Figure 9 shows n_e/v_e for optical glasses in the visible range in the usual form of presentation. Highly refractive glasses tend to have smaller Abbe numbers, i.e., high relative dispersion.

Optical imaging systems require a combination of glasses with different refractive indices. Good imaging quality in normal polychromatic light, however, is not possible unless the dispersion curve has the same partial dispersion in all its sections. In general, the

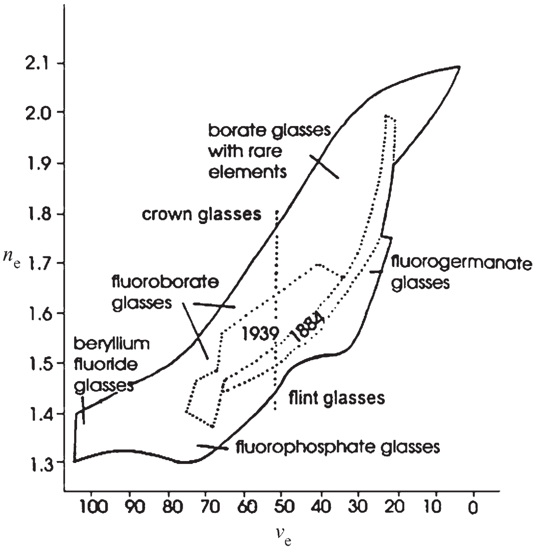


Figure 9
n_e/v_e diagram for optical glasses.

dispersions of two glasses are different, which gives rise to colored rings, an effect called chromatic aberration. If the glasses in a combination of prisms and lenses have different dispersions, a color compensation for two wavelengths is possible in any given case: $(n'_x - n'_y) = c(n''_x - n''_y)$. For prisms, the constant c is the ratio of the small refracting angles, and for lenses it is the ratio of the focal lengths f.

Consequently, the achromatism can be adjusted, for example, to the wavelengths F′ and C′ of the main dispersion (achromats) if the geometrical parameters are suitably chosen. For a pair of lenses, the equation $c = f'/f'' = v'_e/v''_e = (n'_F - n'_C)/(n''_F - n''_C)$ is valid. However, outside the two wavelengths F′ and C′, there is still color separation, the so-called secondary spectrum, and in high-resolution optical systems it is necessary to correct for at least one other wavelength in the middle of the region between F′ and C′. This requires relative partial dispersions to be introduced: $P = \Delta n/(n_{F'} - n_{C'})$, such as $P_{\lambda F'} = (n_\lambda - n_{F'})/(n_{F'} - n_{C'})$. A large series of glasses can be grouped together in a flat curve, which can be approximated by a straight line, the so-called normal straight line, if one of the relative partial dispersion ($\lambda = \text{const.}$) is drawn vs. the Abbe number v_e: $P_{\lambda F'} = a + bv_e + (cv_e^2)$. This empirical relationship of Fig. 10(a) expresses the fact that a change in dispersion in a partial range is in conformity with the change in dispersion in the entire wavelength region under consideration. This case is called normal behavior, or normal partial dispersion.

The secondary spectrum of a lens combination can be described by the relative partial dispersion $P_{\lambda F'}$ as follows: $\Delta S_{\lambda F'} = (P'_{\lambda F'} - P''_{\lambda F'})/(v'_e - v''_e)$, where f is the focal length of the objective. The wavelength λ, and thus n_λ in $P_{\lambda F'}$, now occurs as a variable. As shown in Fig. 10(b), an achromatic color correction for two wavelengths leads to a value of zero because $P'_{\lambda F'}$ and $P''_{\lambda F'}$ are equal to unity for a given wavelength, which results in $\Delta S_{\lambda F'} = 0$. If $P'_{\lambda F'}$ and $P''_{\lambda F'}$ are different from unity but identical in magnitude and sign, the result is a third zero, and thus color correction for a third wavelength (apochromatic color correction). Glass pairs deviating from the normal straight line are said to have an anomalous partial dispersion and they make it possible to design apochromats whose resolving power is better than that of achromats.

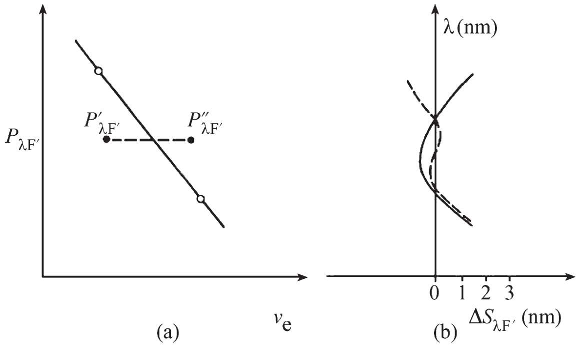


Figure 10
Normal straight (a) $P_{\lambda F'}$ as a function of v_e and (b) representation of the secondary spectrum $\Delta S_{\lambda F'}$ as a function of the wavelength.

Glasses that differ considerably in the position and intensity distribution of their electronic absorption and vibrational spectra, and thus in their structure, have been found suitable for practical use. If, for example, the silicate glasses of a certain group of compositions are correlated to a normal straight line, the deviations will be positive for borate glasses and negative for phosphate glasses. Burckhardt (1984) has shown that glasses with isostructural units, e.g., silicate and germanate glasses, are ruled out for apochromatic combination, because the resonance points corresponding to the electronic and vibrational transitions are shifted in conformity. However, two-lens achromates can be made of such glass pairs. Apochromats require glass pairs in which λ_1 (electronic transition) and λ_2 (vibrational transition) for one glass are either situated both between the corresponding resonance sections of the other glass, or both outside. The resonance sections of the one medium must enclose those of the other. There are also ways to calculate the position of the third zero from structure-related optical parameters.

Bibliography

Ariagada J C, Burckhardt W, Feltz A 1987 The influence of the water content on absorption and dispersion behaviour of calcium phosphate glasses. *J. Non-cryst. Solids.* **91**, 375–85

Barth St, Feltz A 1989 Ion conducting glasses in the system Na_2O–Nb_2O_5–P_2O_5. *J. Non-cryst. Solids* **34**, 41–5

Bell R F, Carnevale A 1981 A structural model for B_2O_3 glass. *Phil. Mag.* **43**, 389–99

Bray P J 1985 Nuclear magnetic resonance studies of glass structure. *J. Non-cryst. Solids* **73**, 19–45

Burckhardt W 1984 The structure of optical materials and the correction of the longitudinal aberration of a two-lens achromate. *Exp. Tech. Phys.* **32**, 193–203

Burckhardt W, Rudolph B 1991 Ionenleitende Gläser-potentielle Festelektrolyte für All-Feststoff-Batterien. *Silicattechnik.* **42**, 275–8

Elliott R J 1997 Evidence for rings in the dynamics of tetrahedral glasses. *J. Non-cryst. Solids* **182**, 1–8

Elliott S R 1988 Frequency-dependent conductivity in ionic glasses: a possible model. *Solid State Ionics* **27**, 131–49

Fanderlik I 1983 *Optical Properties of Glass*. Elsevier, Amsterdam

Feltz A 1977 On the structure of GeS/Ge_2S_2 and GeSe/Ge_2Se_2 glasses and amorphous compounds of Ge_2S_3 and Ge_2Se_3. In: *Proc. 7th Int. Conf. Amorphous Liquid Semiconductors*. Committee by the Centre for Industrial Consultancy and Liaison, University of Edinburgh, UK, Vol. 1

Feltz A 1987 Ionic conducting glasses for batteries and wave guide devices. *J. Non-cryst. Solids* **90**, 545–56

Feltz A 1993 *Amorphous Inorganic Materials and Glasses*. Verlag Chemie, Weinheim, Germany

Feltz A, Büchner P 1987 Na^+ ion conducting glasses in the systems $NaBSiO_4$–Na_2SiO_3, $NaBSiO_4$–Na_4SiO_4 and $NaBSiO_4$–Na_3PO_4. *J. Non-cryst. Solids* **92**, 397–406

Feltz A, Pfaff G 1985 Ordering states in non-crystalline compounds. *J. Non-cryst. Solids* **77/78**, 1137–40

Goodenough J B, Hong H Y P, Kafalas J A 1976 Sodium super ion conductor $Na_{1+3x}Zr_2(PO_4)_{3-3x}(SiO_4)_{3x}$. *Mater. Res. Bull.* **11**, 203

Henn F, Elliott S R, Giuntini J C 1991 Complex permittivity in ionically conducting solids: a hopping model. *J. Non-cryst. Solids* **136**, 60–6

Hunold K, Brückner R 1980 Chemische Diffusion von Na- und Al-Ionen in Natrium-Alumosilicatschmelzen. *Glastrechn. Ber.* **53**, 207–19

Ingram M D, Bunde A, Maass P, Ngai K L 1991 Mixed alkali effects in ionic conductors: a new model and computer simulations. *J. Non-cryst. Solids* **131–3**, 1109–12

Ingram M D, Hunter C C 1984 Na^+ ion conducting glasses. *J. Non-cryst. Solids* **14**, 31–40

Kahnt H 1991 Ionic transport in oxide glasses and frequency dependence of conductivity. *Ber. Bunsenges. Phys. Chem.* **95**, 1021–5

Kahnt H, Schirrmeister F 1987 Ionic diffusion and frequency dependent conductivity in amorphous materials. *J. Non-cryst. Solids* **90**, 421–4

Kaps Ch, Völksch G 1987 On the ion exchange behaviour of the glass $Na_2O{*}2\ SiO_2$ against a melt $(NaNO_3)_{0.9}(TlNO_3)_{0.1}$. *J. Non-cryst. Solids* **91**, 43–51

Lebedev A A 1921 Polymorphism and cooling of glass. *Trudy Gossud. Opt. Inst.* **2**, 57

Levasseur A R, Olazcuaga, Kbala M, Zahir M, Hagenmuller P 1981 Etudes electrique et raman des verres des systemes B_2O_3–M_2O–M_3PO_4 (M: Li, Na). Solid State Ionics **2**, 205–13

Lezikar A V, Simons C J, Moynihan C T 1980 The Debye–Falkenhagen theory of electrical relaxation in glass. *J. Non-cryst. Solids* **40**, 171–88

Liebau F 1985 *Structural Chemistry of Silicates*. Springer, Berlin

Lucovsky G 1979 Spectroscopic evidence for valence-alternation-pair defect states in vitreous SiO_2. *Phil. Mag.* **39**, 513–30

Mackenzie J D 1965 Semiconducting oxide glasses. *Modern Aspects of the Vitreous State* **3**, 126

Susman S, Debecq C J, McMillan J A 1983 NASI glass: a new vitreous electrolyte. *J. Non-cryst. Solids* **9/10**, 667–74

Tomozawa M, Cordaro J F, Singh M 1980 Applicability of weak electrolyte theory to glass. *J. Non-cryst. Solids* **40**, 189–96

Trap H J L, Stevels J M 1959 Physical properties of invert glasses. *Glastechn. Ber.* **32K**, 31–52

Tuller H L, Barsoum M W 1985 Glass solid electrolytes: Past, present and future—the year 2004. *J. Non-cryst. Solids* **3**, 331–50

Tuller H L, Button D P, Uhlmann D R 1980 Fast ion transport in oxide glasses. *J. Non-cryst. Solids* **40**, 93–118

Vogel W 1992 *Glaschemie*. Springer, Berlin

Zachariasen W H 1932 The atomic arrangement in glass. *J. Am. Ceram. Soc.* **54**, 3841–51

A. Feltz
Deutschlandsberg, Austria

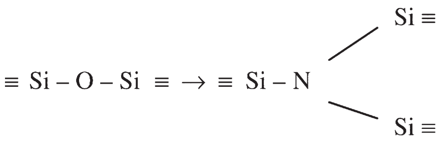
Oxynitride Glasses

Oxynitride glasses are special types of silicate or alumino-silicate glasses in which oxygen atoms in the glass network are partially replaced by nitrogen atoms. Thus, they occur in M–Si–O–N and M–Si–Al–O–N systems where M is a modifying cation such as one of the alkali metals (Li, Na, K); the alkaline earths (Mg, Ca, Ba, Sr or Y), and the rare earth lanthanides.

Following some early work on solubility of nitrogen in silicate glasses, interest in oxynitride glasses grew when it was discovered that silicon nitride based ceramics contained oxynitride glass phases at grain boundaries, as a result of the use of oxide sintering additives. It was found that the composition and volume fraction of these glass phases determine the properties of the material, particularly their high-temperature mechanical behavior. The desire to understand the nature of these intergranular phases resulted in a number of investigations on oxynitride glass formation, structure and properties which have shown oxynitride glasses to possess higher refractoriness, elastic modulus, viscosity and hardness compared to the corresponding oxide glasses as a result of the extra cross-linking provided by nitrogen within the glass structure.

1. *Solubility of Nitrogen in Glasses*

One of the first investigations of nitrogen in silicate glasses (Mulfinger 1966) was concerned with the physical and chemical solubilities of gaseous nitrogen in soda-lime-silica glasses and how this related to inclusions encountered in glass manufacture. The physical solubility of nitrogen in these glasses was found to be extremely low after bubbling nitrogen gas through the glass melt. However, by bubbling ammonia gas through the melt for five hours at a temperature of 1400 °C, the chemical solubility of nitrogen in the glass reached 0.33 wt.% nitrogen, a value 10^5 times higher than that of the physical solubility. It was proposed that nitrogen was substituting for oxygen, under reducing conditions, in the SiO_4 tetrahedra within the silicate network. This leads to a higher than average coordination of non-metal atoms as in:



$$\equiv Si - O - Si \equiv \;\rightarrow\; \equiv Si - N \begin{matrix} Si \equiv \\ \\ Si \equiv \end{matrix}$$

This results in increased cross-linking and, hence, a more rigid glass network.

Elmer and Nordberg (1967) introduced ten times as much nitrogen (3 wt.%) into porous borosilicate glass by heating in ammonia, which resulted in significant increases in annealing temperature, indentation hardness and electrical resistivity, and a marked decrease in the thermal expansion coefficient of these glasses. In this case, the changes observed in the physical properties of the glass resulted from the incorporation of nitrogen into the structure. Devitrification of these glasses could be induced electrolytically and it was noted that incorporation of nitrogen inhibited this devitrification. This was attributed to increased viscosity, due to the presence of (=NH) and (=N–) groups in the glass.

Davies and Meherali (1971) concluded that the solubility of nitrogen in glass melts was chemical rather than physical, and found that severe reducing conditions had to be imposed in order to dissolve significant amounts of nitrogen into glass melts. The solubility of nitrogen increased with increasing basicity, indicating that bridging rather than non-bridging oxygen atoms were involved in the dissolution reaction.

An investigation of the solubility of nitrogen in $CaO–SiO_2–Al_2O_3$ metallurgical slags, (Dancy and Janssen 1976) compared physical and chemical methods of dissolving nitrogen in these melts. Under 1 atm. of nitrogen, an equilibrium solubility of 0.25–2.5 wt.% nitrogen was achieved after 24 h. By contrast, when Si_3N_4 was added to the melt, again under a nitrogen atmosphere, nitrogen incorporation was very rapid and reached significantly higher levels (4 wt.%). It was suggested that most silicate melts should not be significantly reducing to dissolve N_2 or NH_3 to any great extent and that it was necessary to dissolve nitrogen from nitrides such as silicon nitride.

Jack (1976) noted that the local atomic arrangement in Si_3N_4 is a tetrahedral array of nitrogen atoms around a central silicon atom, similar to the SiO_4 tetrahedron within the network of silicate glasses. The lengths of Si–N, Si–O and Al–O bonds are similar and it was proposed therefore that nitrogen could be easily incorporated into the network of silicate and alumino-silicate glasses.

2. Glass Formation in M–Si–Al–O–N Systems and Their Representation

Many investigations have been carried out on glass formation and properties in several M–Si–O–N and M–Si–Al–O–N systems (where M = Modifying Cation) and these are listed in Table 1.

Preparation of oxynitride glasses (Hampshire *et al.* 1994) involves wet ball milling of appropriate powders—silica, alumina, the modifying oxide plus silicon nitride or aluminum nitride—in isopropyl alcohol using sialon milling media, followed by evaporation of the alcohol. Typically, large batches (50–60 g) are melted in boron nitride lined graphite crucibles under 0.1 MPa nitrogen pressure at 1700–1750 °C for 1 h in a vertical tube furnace, after which the melt is poured into a preheated graphite mould at 850 °C. The glass is annealed at this temperature for one hour to remove stresses and slowly cooled.

Convenient methods of representing both Si–Al–O–N and M–Si–Al–O–N systems involve the concept of reciprocal salt pairs. The four-component Si–Al–O–N system is a square plane which has, as components, the oxides and nitrides of silicon and aluminum. The introduction of a further cation gives a five-component system represented by Jänecke's triangular prism shown in Fig. 1 for the Y–Si–Al–O–N system in which the boundaries of the complete glass forming region are outlined.

The concentrations of all components are expressed in equivalents instead of atoms or gram-atoms. One equivalent of any element always reacts with one equivalent of any other element or species. For a system containing two types of cations, A and B with valencies of v_A and v_B, respectively, then:

$$\text{equivalent concentration of A} = (v_A[A])/(v_A[A] + v_B[B])$$

where $[A]$ and $[B]$ are, respectively, the atomic concentrations of A and B.

If the system also contains two types of anions, C and D with valencies v_C and v_D, respectively, then:

$$\text{equivalent concentration of C} = (v_C[C])/(v_C[C] + v_D[D])$$

where $[C]$ and $[D]$ are, respectively, the atomic concentrations of C and D.

It is convenient to let the right hand top corner of the basal square represent one mole of Si_3N_4. In going from right to left, 3 Si^{4+} is gradually replaced by $4Al^{3+}$ and, from top to bottom, $4N^{3-}$ is replaced by $6O^{2-}$, so that the corners of the square are Si_3N_4,

Table 1
Studies of glass formation and properties of oxynitride glasses in different M–Si–O–N and M–Si–Al–O–N systems.

Modifying cation (M) studied in M–Si–O–N and M–Si–Al–O–N systems	References
Ba	Tredway and Risbud (1983)
Ca	Desmaison-Brut *et al.* (1988), Drew *et al.* (1981), Loehmann (1983), Sakka *et al.* (1983)
Ce	Ohashi and Hampshire (1991)
Li	Messier and Gleisner (1988), Reau *et al.* (1993)
Ln = La, Ce, Nd, Sm, Eu, Dy, Gd, Ho, Er, Yb	Menke *et al.* (2000), Murakami and Yamamoto (1994), Ohashi *et al.* (1995), Ramesh *et al.* (1997), Rocherulle *et al.* (1989)
Mg	Drew *et al.* (1981), Hampshire *et al.* (1985), Leng-Ward *et al.* (1986), Tredway and Risbud (1984)
Nd	Drew *et al.* (1981), Hampshire *et al.* (1985), Ramesh *et al.* (1996)
Sc	Tredway and Loehmann (1985)
Y	Drew *et al.* (1981), Hampshire *et al.* (1994), Lemercier *et al.* (1996), Lemercier *et al.* (1997), Leng-Ward and Lewis (1985), Loehmann (1983), Messier (1987), O'Meara *et al.* (1992)
Mixed cations	Pomeroy *et al.* (2003), Weldon *et al.* (1996)

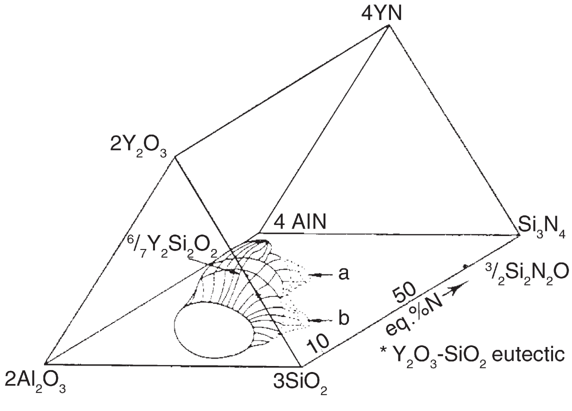


Figure 1
Three-dimensional representation of the complete glass forming region in the Y–Si–Al–O–N Jänecke prism (after Drew *et al.* 1981, 1983).

Si_3O_6, Al_4O_6 and Al_4N_4. By moving into the prism, the third "cation" also needs to be taken into account. Any point in the prism represents a combination of 12 positive and 12 negative valencies. The concentrations can be expressed as percentages so that:

$$\text{equivalent percentage (e/o) of Si} = (4[Si]) \times 100/(4[Si] + 3[Al] + v_M[M])$$

where $[Si]$, $[Al]$ and $[M]$ are, respectively, the atomic concentrations of Si, Al and the third cation M (in this case Y) in any composition and v_M is the valency of M. Any vertical plane has a constant nitrogen:oxygen ratio on which:

$$\text{equivalent percentage (e/o) of nitrogen} = (3[N]) \times 100/(2[O] + 3[N])$$

where $[O]$ and $[N]$ are, respectively, the atomic concentrations of oxygen and nitrogen within any composition.

This representation was adopted by Drew *et al.* (1981, 1983) and Hampshire *et al.* (1985, 1994) to describe the limits of glass formation for Mg– and Y–Si–Al–O–N systems. The limits of the metal alumino-silicate glass regions are plotted on the oxide face of the prism and the glass forming region is seen to expand initially as nitrogen is introduced and then diminishes above ~10 equivalent per cent (e/o) N until the solubility limit for nitrogen is exceeded. Depending on the particular system, it was found that a limit of 17–25 e/o of oxygen could be replaced by nitrogen.

In M–Si–O–N systems, much smaller glass-forming regions are observed as can be seen from the small areas marked a and b in Fig. 1. This confirms the ability of Al_2O_3 to extend the range of glass formation in both silicate and oxynitride systems. The miscibility gap in various silicate binary systems extends into the equivalent M–Si–O–N system with resulting phase separation in most of the glasses, as found by Ohashi and Hampshire (1991) in the Ce–Si–O–N system and shown in Fig. 2.

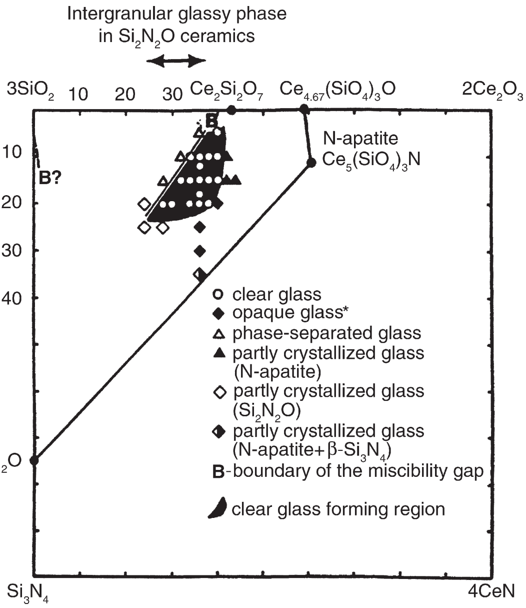

Intergranular glassy phase
in Si_2N_2O ceramics
$3SiO_2$
10
20
30
$Ce_2Si_2O_7$
$Ce_{4.67}(SiO_4)_3O$
$2Ce_2O_3$
N-apatite
$Ce_5(SiO_4)_3N$
B
B?
10
20
30
40
Si_2N_2O
Si_3N_4
4CeN
o clear glass
◆ opaque glass*
△ phase-separated glass
▲ partly crystallized glass (N-apatite)
◇ partly crystallized glass (Si_2N_2O)
◆ partly crystallized glass (N-apatite+β-Si_3N_4)
B-boundary of the miscibility gap
clear glass forming region

Figure 2
Glass formation in the Ce–Si–O–N system (after Ohashi and Hampshire 1991).

3. *Effects of Nitrogen on Properties and Structure of Oxynitride Glasses*

Early work on the correlations between nitrogen incorporation into oxynitride glasses and changes in their physical properties was reported by Loehmann (1980, 1983). Glass transition temperature (T_g), microhardness and fracture toughness increased with increasing nitrogen content, while the thermal expansion coefficient decreased. IR spectroscopic analysis indicated that the incorporated nitrogen became bonded to silicon in the glass network and by substitution for oxygen, produced a more tightly and highly linked structure. These reports did not take into account changes in the Al or the modifying cation concentration and so these property changes could not be attributed solely to incorporation of nitrogen.

The first systematic studies on the effect of replacing oxygen by nitrogen in oxynitride glasses with fixed cation compositions were reported by Drew *et al.* (1981, 1983) and Hampshire *et al.* (1984, 1985). For all Mg–, Ca–, Y–, and Nd–S–Al–O–N glasses with a constant cation ratio, incorporation of nitrogen resulted in increases in glass transition temperature (T_g), shown in Fig. 3 and also viscosity, resistance to devitrification, refractive index, dielectric constant and a.c. conductivity. Figure 4 shows that viscosity increases by more than two orders of magnitude simply by replacing 18 e/o oxygen by nitrogen.

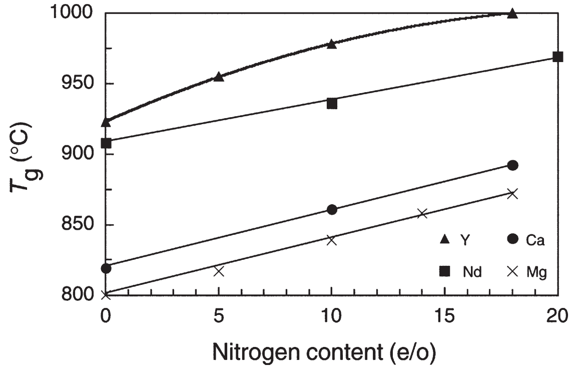


Figure 3
Effect of nitrogen on glass transition temperature (T_g) of Mg–, Ca–, Nd–, and Y–Si–Al–O–N glasses (after Drew *et al.* 1981).

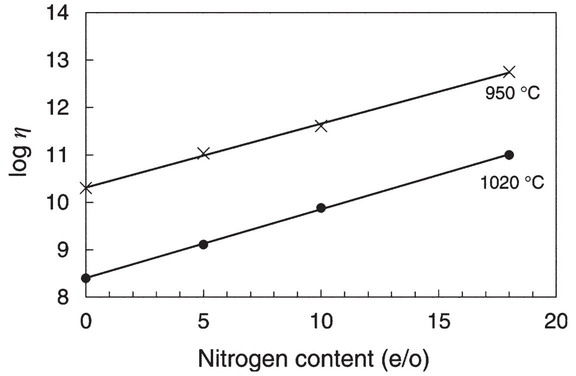


Figure 4
Effect of nitrogen on viscosity of Y–Si–Al–O–N glasses (after Hampshire *et al.* 1984).

In the Y–Si–Al–O–N system (Hampshire *et al.* 1994), it was found that T_g, viscosity; microhardness; Young's and shear moduli all increase systematically with increasing nitrogen:oxygen ratio for different series of glasses. This same study also investigated the effect of fixed Si:Al and Y:Al ratios on properties of glasses with constant O:N ratio. As Si content increases, T_g and viscosity increase while elastic moduli; hardness and thermal expansion coefficient decrease. With increasing Al:Y ratio, elastic moduli and thermal expansion coefficient decrease while the T_g and viscosity decrease to a minimum (at 16 e/o Al) and then increase with further increase in Al content. It appears that changes in T_g and viscosity may be governed by simultaneous mechanisms with opposing effects:

(i) the enhancement of the cross-linking of the glass network, so reducing the number of nonbridging anions, which dominates at high Al:Y ratios when 4 coordinated Al is prevalent.

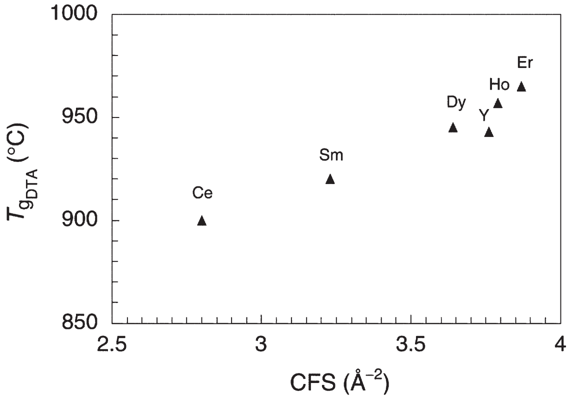


Figure 5
Effect of rare earth cation on glass transition temperature (T_g) of Ln–Si–Al–O–N glasses; Ln = Ce, Sm, Dy, Ho, Er (after Ramesh *et al.* 1997).

(ii) the decrease in network compactness due to the increase in 5 or 6 coordinated Al atoms at lower Al:Y ratios.

Properties of various Ln–Si–Al–O–N glasses have been reported (Menke *et al.* 2000, Murakami and Yamamoto 1994; Ohashi *et al.* 1995; Ramesh *et al.* 1997). A decrease in the lanthanide ion radius, and hence an increase in the cation field strength, results in increases in T_g shown in Fig. 5, and also viscosity, Young's modulus and hardness while molar volume decreases.

In Ln–Si–Al–O–N glasses with higher Ln and Al content, Young's modulus and hardness increase linearly with cation field strength (decreasing cation radius) but properties such as thermal expansion coefficient and glass transition temperature do not show this linearity and there is a discontinuity in all properties within the lanthanide series observed at Gd (Menke *et al.* 2000).

3.1 Nitrogen Coordination in Glasses

The resulting improvements in glass properties by substitution of nitrogen for oxygen has usually been attributed to an increase in the cross-linking of the silicate network due to the replacement of a two-coordinated bridging oxygen atom, by a bridging nitrogen atom coordinated by three silicon ions. Extensive studies on the coordination of nitrogen in oxynitride glasses using Fourier transform infrared spectroscopy (FTIR), x-ray photoelectron spectroscopy (XPS), nuclear magnetic resonance (NMR) and neutron diffraction have been carried out by Brow and Pantano (1984), Hater *et al.* (1989), and Jin *et al.* (1993). Rouxel *et al.* (1990) also used Raman spectroscopy. The structural features are as follows:

(a) Nitrogen is present in the structural network as Si–N bonds, as evidenced by the shifting of the position of the Si–O–Si stretching peak towards that of Si–N. If nitrogen exists only as precipitated Si_3N_4, the position of the Si–O–Si peak would not be expected to change. The preference is for Si–N bonding over Al–N as indicated by ^{29}Si NMR.

(b) Nitrogen is present in threefold coordination (suggested by XPS studies and consistent with systematic changes in the physical and chemical properties of the glasses). However, some nitrogen atoms are bonded to only two Si atoms, or even one, instead of three, as in:

$$\text{(a)}\ \equiv \mathrm{Si} - \mathrm{N}^{-} - \mathrm{Si} \equiv$$

$$\text{(b)}\ \equiv \mathrm{Si} - \mathrm{N}^{2-}$$

This suggests that nonbridging nitrogen atoms may also be present. The local charge on the nonbridging nitrogen ions is balanced by the presence of interstitial metal ions in their vicinity. In the case of silicate glasses, nonbridging oxygen atoms replace bridging oxygen atoms at high modifier contents. For (a) above, while the N atom links two silicon atoms rather than three, it is still effectively a "bridging" ion.

(c) The glass network contains $Si(O_4)$, $Si(O_3N)$ and $Si(O_2N_2)$ tetrahedral structural groups identified by ^{29}Si NMR–MAS.

4. *Crystallization of Oxynitride Glasses to Form Glass-ceramics*

As with other silicate glasses, oxynitride glasses may be heat treated at the appropriate temperature to crystallize as glass-ceramics (Leng-Ward and Lewis 1990, Thompson 1992, Hampshire 1993). The crystalline phases formed depend on both the composition of the parent glass and the heat-treatment process. The conventional process to produce a glass-ceramic involves two steps: a low temperature heat treatment of glasses to induce nucleation followed by heating to a second higher temperature to allow crystal growth of the nuclei. Many glasses require the addition of a nucleating agent to promote the crystallization process (Tredway and Risbud 1984) but oxynitride glasses appear to be self-nucleating because nucleation occurs heterogeneously on FeSi particles as found by Dinger *et al.* (1988) and Besson *et al.* (1993), in Y_2O_3–SiO_2–AlN glasses.

Many studies of crystallization of oxynitride glasses have concentrated on the Y–Si–Al–O–N system. The glass-ceramic transformations in a glass of composition (in e/o) 28Y:56Si:16Al:830:17N were studied by Ramesh *et al.* (1998) using both classical and differential thermal analysis techniques and these two methods were found to be in close agreement. Optimum nucleation and crystallization temperatures were determined in relation to the glass transition temperature. The major crystalline phases present are mixtures of different forms of yttrium disilicate and silicon oxynitride. Bulk nucleation was observed to be the dominant nucleation mechanism. The activation energy for the crystallization process was found to be 834 KJ mol^{-1}.

Leng-Ward and Lewis (1985) studied the crystallization at 1250 °C of a series of glasses of constant cation composition (in e/o) 26Y:42Si:32Al with zero to 30 e/o nitrogen. As nitrogen content increases, gradual replacement of yttrium disilicate phase by yttrium aluminum garnet (YAG) was observed and nitrogen was mainly incorporated into Si_2N_2O.

Hampshire *et al.* (1994) studied the crystallization behavior of a glass (in e/o) 35Y:45Si:20Al containing 23 e/o nitrogen and found that B-phase (Y_2SiAlO_2N), Iw-phase ($Y_2Si_3Al_2(O,N)_{10}$) and wollastonite ($YSiO_2N$) formed at temperatures below 1200 °C while α-yttrium disilicate ($Y_2Si_2O_7$), apatite ($Y_2Si_3O_{12}N$) and YAG ($Y_3Al_2O_{12}$) formed at higher temperatures. At relatively low heat treatment temperatures of ~950–1100 °C, the nucleation and growth of N–wollastonite ($YSiO_2N$) and the intermediate phases B and Iw are kinetically favored over that of the more stable equilibrium phases YAG and Si_2N_2O. In a study of the initial stages of crystallization of this same glass (Besson *et al.* 1997), it was found that the creep rate is higher than for the parent glass, since the residual glass following nucleation has a lower viscosity due to a decrease in yttrium and increase in impurity cations. Following complete crystallization, the creep rate was very low.

Further studies on the crystallization of B and Iw phases in these Y–Si–Al–O–N glasses have been reported (Lemercier *et al.* 1997, Gonon *et al.* 2000, Young *et al.* 2000, MacLaren *et al.* 2001, Diaz and Hampshire 2002).

5. *Effect of Nitrogen Content, Surface Condition, and Heat Treatment on Fracture of Oxynitride Glasses*

The fracture behavior of glasses of composition (in e/o) 35Y:45Si:20Al with 0 e/o N and 17 e/o N was assessed (Lemercier *et al.* 2002) by means of flexural strength measurements and using a fractographic approach. Heat treatments at T_g–20 °C and T_g + 20 °C were carried out on polished flexural specimens in order to round out the crack tips produced during the surface-finishing step.

Without any heat treatment (just after polishing), the oxide glass exhibits a slightly higher flexural strength of 191 MPa compared with 162 MPa for the oxynitride glass. For the oxide glass, a heat treatment at T_g–20 °C results in an increase in strength to 295 MPa, which remains essentially constant after a treatment at T_g + 20 °C (310 MPa). The strength of the oxynitride glass after treatment at T_g–20 °C (344 MPa) is twice the strength measured on the nontreated glass and, after treatment at T_g + 20 °C, is three times higher (479 MPa), resulting in a significant difference in fracture strength between the two glasses, suggesting that nitrogen strengthens the glass network.

A double stage glass-ceramic heat treatment (960 °C for 1 h and 1050 °C for 5 h), which allows crystal growth to form B-phase, was carried out on the 17 e/o N glass samples. A strength of 549 MPa was measured for the glass-ceramic.

Based on a statistical distribution of location of fracture origins, it appears that:

(i) Without any heat treatment, fracture occurs from defects present on both the plane surface and the chamfer. The surface finish does not affect this observed behavior.

(ii) After a heat treatment at T_g–20 °C, fracture occurs mainly from the plane surface.

(iii) After a heat treatment at T_g + 20 °C, the fracture origins were in the bulk of the material, regardless of composition, size or surface finish.

6. *Potential Applications*

Specialty or "new" glasses with novel functions are being used in modern industrial sectors such as optoelectronics, microelectronics, communications technologies, biomedical devices and niche areas of the automotive and architectural sectors. Oxynitride glasses are one addition to the range of "new" glasses and, in view of their durability, higher refractoriness, tailored thermal properties, etc., areas of potential application are in: passive coatings on electronic substrates; novel glaze systems for refractory protection; novel glass decoration enamels; high temperature joining of ceramics and of ceramics to metals and higher temperature range glass-ceramics for structural applications.

7. *Concluding Remarks*

Oxynitride glass formation occurs in a number of M–Si–Al–O–N systems and up to 30 e/o nitrogen can be accommodated within the glass. As nitrogen increases, glass transition temperatures, elastic modulus, viscosity and hardness increase while thermal expansion coefficient decreases. The effect of nitrogen on properties is much greater than the effect of cations. For a constant cation composition, viscosity increases by more than two orders of magnitude as 18 e/o oxygen is replaced by nitrogen. For rare earth sialon glasses of constant composition; viscosity;

Young's modulus and T_g increase linearly with decreasing ionic radius. Spectroscopic studies have identified the structural features of glasses and the role of nitrogen.

For two glasses in the YSiAlON system with nitrogen contents of zero and 17 e/o N, appropriate heat treatments at a temperature close to T_g promote the blunting of surface cracks by viscous flow, which is improved by increasing the heat treatment temperature slightly over T_g. This results in a flexural strength of 480 MPa for the oxynitride glass as a result of a change in the location of the failure origin from the surface to the bulk of the material.

Many studies on crystallization of oxynitride glasses have been carried out. Some of these have identified suitable two stage heat treatments for nucleation and growth of crystal phases to form glass ceramics with significant increases in strength over the parent glass.

Bibliography

Besson J-L, Billiers D, Rouxel T, Goursat P, Flynn R, Hampshire S 1993 Crystallization and properties of a Si–Y–Al–O–N glass-ceramic. *J. Am. Ceram. Soc.* **76**, C2103–C2105

Besson J-L, Lemercier H, Rouxel T, Troillard G 1997 Yttrium sialon glasses: nucleation and crystallisation of $Y_{35}Si_{45}Al_{20}O_{83}N_{17}$. *J. Non-Cryst. Sol.* **211**, 1–21

Brow R K, Pantano C G 1984 Nitrogen coordination in oxynitride glasses. *J. Am. Ceram. Soc.* **67**, C72–C74

Dancy E A, Janssen D 1976 Dissolution of nitrogen in metallurgical slags. *Can. Metall. Q.* **15**, 103–10

Davies M W, Meherali S G 1971 Equilibrium solubility of nitrogen in aluminosilicate melts. *Metall. Trans.* **2**, 2729–33

Desmaison-Brut M, Desmaison J, Verdier P 1988 Oxidation behaviour of an oxynitride glass in the system Ca–Si–Al–O–N. *J. Non-Cryst. Sol.* **105**, 323–9

Diaz A, Hampshire S 2002 Crystallisation of M–SiAlON glasses to Iw-phase glass-ceramics: preparation and characterization. *J. Mater. Sci.* **37**, 723–30

Dinger T R, Rai R S, Thomas G 1988 Crystallization behaviour of a glass in the Y_2O_3–SiO_2–AlN system. *J. Am. Ceram. Soc.* **71**, 236–44

Drew R A L, Hampshire S, Jack K H 1981 Nitrogen glasses. *Proc. Brit. Ceram. Soc.* **31**, 119–32

Drew R A L, Hampshire S, Jack K H 1983 The preparation and properties of oxynitride glasses. In: Riley F L (ed.) *Progress in Nitrogen Ceramics, Proc. NATO Advanced Study Institute.* Martinius Nijhoff, The Hague, pp. 323–30

Elmer T H, Nordberg M E 1967 Effect of nitriding on electrolysis and devitrification of high-silica glasses. *J. Am. Ceram. Soc.* **50**, 275–9

Gonon M F, Descamps J C, Cambier F, Thompson D P 2000 Determination and refinement of the crystal structure of M_2SiAlO_5N (M = Y, Er, Yb). *Ceram. Internat.* **26**, 105–11

Hampshire S 1993 Oxynitride glasses and glass-ceramics. In: Chen I W, Becher P F, Mitomo M, Petzow G, Yen T-S (eds.) *Silicon Nitride Ceramics—Scientific and Technological Advances. Mater. Res. Soc. Symp. Proc.*, Vol. 287, pp. 93–100

Hampshire S, Drew R A L, Jack K H 1984 Viscosities, Glass Transition Temperatures and Microhardness of Y–Si–Al–O–N glasses. *J. Am. Ceram. Soc.* **67**, C46–C47

Hampshire S, Drew R A L, Jack K H 1985 Oxynitride glasses. *Phys. Chem. Glass.* **26**, 182–6

Hampshire S, Nestor E, Flynn R, Besson J-L, Rouxel T, Lemercier H, Goursat P, Sebai M, Thompson D P, Liddell K 1994 Yttrium oxynitride glasses: properties and potential for crystallisation to glass-ceramics. *J. Euro. Ceram. Soc.* **14**, 261–73

Hater W, Warmuth W M, Frischat G H 1989 ^{29}Si MAS NMR studies of alkali silicate oxynitride glasses. *Glastechn. Ber.* **62**, 328–35

Jack K H 1976 Review: sialons and related nitrogen ceramics. *J. Mater. Sci.* **11**, 1135–58

Jin J, Yoko T, Miyaji F, Sakka S, Fukunaga T, Misawa M 1993 Neutron diffraction study on the structure of NaSiON oxynitride glass. *J. Am. Ceram. Soc.* **76**, 630–4

Lemercier H, Rouxel T, Fargeot D, Besson J-L, Pirou B 1996 Yttrium SiAlON glasses: structure and mechanical properties-elasticity and viscosity. *J. Non-Cryst. Sol.* **201**, 128–45

Lemercier H, Ramesh R, Besson J-L, Liddell K, Thompson D P, Hampshire S 1997 Preparation of pure B-phase glass-ceramic in the yttrium-sialon system. *Key. Eng. Mater.* **132–136**, 814–17

Lemercier M, Servidio S, Hampshire S 2002 *Phil. Mag.*, (submitted)

Leng-Ward G, Lewis M H 1985 Crystallisation in Y–Si–Al–O–N glasses. *J. Mater. Sci. Eng.* **71**, 101–11

Leng-Ward G, Lewis M H 1990 Oxynitride glasses and their glass-ceramic derivatives. In: Lewis M H (ed.) *Glasses and Glass–Ceramics.* Chapman and Hall, London

Leng-Ward G, Lewis M H, Wild S 1986 Crystallisation of 3M/4X Mg–Si–Al–O–N melts. *J. Mater. Sci.* **21**, 1647–53

Loehmann R E 1980 Oxynitride glasses. *J. Non-Cryst. Sol.* **42**, 433

Loehmann R E 1983 Preparation and properties of oxynitride glasses. *J. Non-Cryst. Sol.* **56**, 123–34

MacLaren I, Falk L K L, Diaz A, Hampshire S 2001 Effect of composition and crystallization temperature on microstructure of Y– and Er–SiAlON I_w–phase glass-ceramics. *J. Am. Ceram. Soc.* **84**, 1601–8

Menke Y, Peltier-Baron V, Hampshire S 2000 Effect of rare-earth cation on properties of SiAlON glasses. *J. Non-cryst. Sol.* **276**, 145–50

Messier D R 1987 Preparation and properties of Y–Si–Al–O–N glasses. *Int. J. High Tech. Ceram.* **3**, 33–41

Messier D R, Gleisner R P 1988 Preparation and characterisation of Li–Si–Al–O–N glasses. *J. Am. Ceram. Soc.* **71**, 422–5

Mulfinger H O 1966 Physical and chemical solubility of nitrogen in glass melts. *J. Am. Ceram. Soc.* **49**, 462–7

Murakami Y, Yamamoto H 1994 Properties of oxynitride glasses in the Ln–Si–Al–O–N systems (Ln = rare-earth). *J. Ceram. Soc. Jpn.* **102**, 231–6

Ohashi M, Hampshire S 1991 Formation of Ce–Si–O–N glasses. *J. Am. Ceram. Soc.* **74**, 2018–20

Ohashi M, Nakamura K, Hirao K, Kanzaki S, Hamshire S 1995 Formation and properties of Ln–Si–O–N glasses. *J. Am. Ceram. Soc.* **78**, 71–6

O'Meara C, Dunlop G L, Pompe P 1992 Formation, crystallisation and oxidation of selected glasses in the Y–Si–Al–O–N system. *J. Euro. Ceram. Soc.* **8**, 161–70

Pomeroy M J, Mulcahy C, Hampshire S 2003 Independent effects of nitrogen substitution for oxygen and magnesium substitution by yttrium on the properties of Mg–Y–Si–Al–O–N glasses. *J. Am. Ceram. Soc.* **86**, (in press)

Ramesh R, Nestor E, Pomeroy M J, Hampshire S, Liddell K, Thompson D P 1996 Potential of NdSiAlON glasses for crystallisation to glass-ceramics. *J. Non-Cryst. Sol.* **196**, 320–5

Ramesh R, Nestor E, Pomeroy M J, Hampshire S 1997 Formation of Ln–Si–Al–O–N glasses and their properties. *J. Euro. Ceram. Soc.* **17**, 1933–9

Ramesh R, Nestor E, Pomeroy M J, Hampshire S 1998 Classical and DTA studies of the glass-ceramic transformation in a YSiAlON glass. *J. Am. Ceram. Soc.* **81**, 1285–97

Reau J M, Kahnt H, Rocherulle J, Verdier P, Laurent Y 1993 The influence of nitrogen on the mobility of lithium in oxynitride glasses of the Li–Si–Al–O–N system. *J. Non-Cryst. Sol.* **155**, 185–8

Rocherulle J, Verdier P, Laurent Y 1989 Preparation and properties of gadolinium oxide and oxynitride glasses. *Mater. Sci. Eng.* **82**, 265–8

Rouxel T, Besson J-L, Rzepka E, Goursat P 1990 Raman spectra of SiYAlON Glasses and Ceramics. *J. Non-Cryst. Sol.* **122**, 298–304

Sakka S, Kamiya K, Yoko T 1983 Preparation and properties of Ca–Al–Si–O–N oxynitride glasses. *J. Non-Cryst. Sol.* **56**, 147–55

Thompson D P 1992 Oxynitride glasses. In: Cable M, Parker J M (eds.) *High Performance Glasses*. Blackie and Sons, Glasgow Chap. 5

Tredway W K, Loehmann R E 1985 Scandium-containing oxynitride glasses. *J. Am. Ceram. Soc.* **68**, C131–C133

Tredway W K, Risbud S H 1983 Preparation and controlled crystallisation of Si–Ba–Al–O–N oxynitride glass. *J. Non-Cryst. Sol.* **56**, 135–40

Tredway W K, Risbud S H 1984 Influence of atmospheres and TiO_2 nucleant on the crystallisation of Mg–SiAlON glasses. *J. Mater. Sci. Lett.* **4**, 31–3

Weldon L, Pomeroy M J, Hampshire S 1996 Glasses and Ceramics in the rare-earth sialon systems. *Key. Eng. Mater.* **118–119**, 241–8

Young W, Falk L K L, Lemercier H, Peltier-Baron V, Menke Y, Hampshire S 2000 *J. Non-Cryst. Solids* **270**, 6–19

S. Hampshire
University of Limerick, Ireland

P

Paper: Porosity

The porosity of paper determines a number of the final properties such as absorptivity, opacity, strength, and ink–paper interactions. The porosity and the pore size distribution can be measured by a number of techniques, but most of these techniques have some difficulties. The porosity is determined by the packing characteristics of fibers, fines and pigments, chemical additives, and processing conditions. The porosity of paper is not uniform, but can vary in both the thickness direction and within the plane. For certain grades of paper, a coating is applied to the top surface of the paper to modify the porosity of the top surface.

The porosity of paper is especially important for some paper products that are required to absorb fluids such as tissue- and towel-type grades. Printing is strongly influenced by the porosity of the surface layer (Bureau 1976). The barrier properties and strength properties are influenced by the paper porosity. The porosity of the top surface of paper surfaces also influences the performance of ink jet printing. Therefore, the porosity of paper is an important topic of study.

1. Measurement Techniques

There are a number of methods to characterize the porosity or void fraction of paper, as described by others (Murakami and Imamura 1984). These methods can be classified in various ways and give different levels of detail. In simple terms, porosity is the volume of the voids divided by the total volume of the sample. However, to measure either of these quantities is not simple because of the complex nature of paper. The volume of the sample can be estimated from a caliper method along with the dimensions of the sample. However, the roughness of paper and the compressibility of the sample will generate some uncertainty in this measurement. If the density of the fiber and pigment component of the sheet can be estimated, then the porosity can be estimated from the weight of the sample.

Other techniques involve the displacement of the voids with a gas or liquid, and measuring the amount of fluid that is taken up by the sample. With these techniques, the sample is placed in a vacuum, and the amount of fluid that fills the voids is measured. Helium is considered a desirable gas because of its small molecular size. Other wetting fluids such as water or silicon oils can be used to obtain a good estimate, but questions arise about the ability of some fluids to penetrate all of the voids and the modification of the pore structure by the probe fluid.

The mercury intrusion method has found widespread acceptance, and is of special interest. This method gives the void volume and the pore size distribution by measuring the amount of intrusion as a function of the applied pressure. Mercury is a nonwetting fluid and, therefore, large pressures are required to force mercury into the small pores. By measuring the intrusion volume as a function of pressure, and assuming a contact angle, the volume of voids at various radii is obtained. Figure 1 is an example of mercury intrusion data for a coating on plastic film and a coating on paper. Note that the coating on paper has two peaks in the pore size distribution, one at 2.5 μm that is associated with the paper and one at 0.06 μm associated with the coating layer.

The coating on plastic film only has small pores and a peak at about 0.1 μm. This peak is larger than the coating on paper because it has a larger pigment size. At high pressure, a number of corrections must be made to account for the compression of mercury, expansion of the sample chamber, and compression of the solid fraction. As the pressure is reduced in this method, the volume–pressure curve normally has a hysteresis. Pits or large pores within the structure that are only connected to the outside surface by smaller pore cause this hysteresis.

A common experimental method in industry to characterize porosity is by air permeability. A few standard test methods and devices such as the "Gurley" or the "Sheffield" measure the rate of air flow through a sample for a given pressure drop across the thickness. High porosity corresponds to high airflow rates in general. However, this measurement is not an actual measurement of porosity,

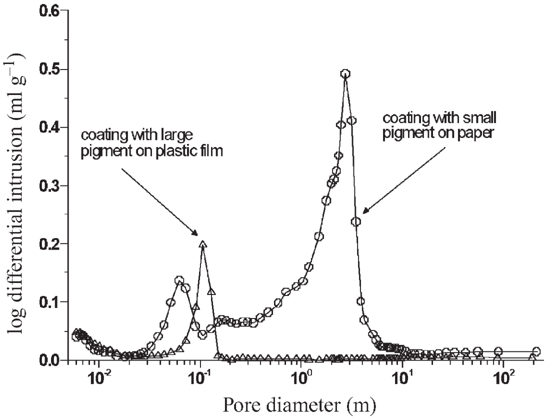


Figure 1
Mercury penetration results for a coating on plastic film and on paper. The pigment diameters on plastic and paper are 0.5 and 0.1 μm, respectively.

because the air flow rate is a function of the paper thickness, the connectivity of the pore space, and the size of the pores. This test is good for a comparison and quality control, but it does not give actual porosity values. For flow through a porous media, a common expression used for paper is the Carman–Kozey equation, given as

$$v = \frac{\Delta P}{k\mu L S^2 (1-\varepsilon)^2 / \varepsilon^3}$$

where v is the average velocity of the fluid, ΔP is the pressure difference across the sample, k is a constant equal to 4.17 for most situations, μ is the fluid viscosity, L is the thickness of the sample, S is the specific surface area (area/volume), and ε is the void fraction. Of all the parameters, the void fraction is the most influential on the flow rate, but the sample thickness and specific surface area are not as easy to obtain. However, the flow rate measurement is a quick and convenient method to compare samples and obtain some idea of porosity.

One interesting technique combines the air permeability with a liquid displacement type idea, and is described in more detail by Murakami and Imamura (1984). In this method, the sample is saturated with some wetting fluid. An increasing pressure difference is applied across the sample, and the rate of airflow is measured as a function of pressure. When the pressure difference that matches the capillary pressure holding the fluid in the capillaries is reached, the capillary will empty, and the air flow rate will increase. Therefore, the porosity and the pore size distribution are obtained with this method.

Other methods are possible to characterize the void volume and the pore size distribution. Cross-sectional examination with a scanning electron microscope and computer analysis may give some detailed information (Allem and Uesaka 1999). Nitrogen adsorption will characterize the specific surface area of a sample. Light-scattering measurements can help reveal the porosity and pore size if assumptions are made with regard to the materials index of refraction.

2. Process Parameters that Influence Porosity

In general terms, the porosity of paper is determined by the packing of fibers and pigments in the sheet during processing. For example, decreasing the amount of fine material should tend to increase porosity. A number of other factors influence the porosity of paper such as chemical additives used for retention, the preconditioning of the fibers through mechanical action, drying conditions, and calendering conditions. The pore size distribution is often taken to be a log-normal distribution, as shown by Niskanen *et al.* (1998).

Mechanical agitation, "beating" or "refining," is one method to modify the paper fiber and the final paper porosity. Wood fibers change their physical characteristics with mechanical agitation or "beating"; microfibers are released from the fiber surface increasing the ability of fibers to hydrogen bond with other fibers. In addition, fine fiber fragments are generated. Refining normally decreases the porosity and decreases the pore size distribution (Murakami and Imamura 1984). The compression of fiber mats during drainage is described by Sayegh and Gonzalez (1995).

After paper is formed with a filtration step, the wet mat is pressed to remove more water and dried with heat. During the drying step, water leaves the structure, and surface tension forces of the water pull the components of the paper together. The dimensions of the mat decrease, and the amount of shrinkage determines to some extent the porosity of the paper (see *Paper: Structure*). The shrinkage is a complex process, and is linked to other issues such as cockle and curl. Rance (1980) gives a detail review of the consolidation of the paper web and shrinkage during drying.

Calendering of paper is often carried out to increase the gloss properties of the sheet. Special coated grades of paper are calendered to produce high-gloss properties that produce desirable print properties and optical properties. Calendering involves the application of a compression force in a nip with a short duration. The ideal case is when the surface roughness of the paper is reduced while retaining the bulk properties. Often, the entire sheet is compressed, and the porosity is reduced with calendering. The influence of calendering on paper porosity is discussed in Kershaw (1980).

With computer simulations and mathematical statistics, the structure of paper and its pore space can be characterized and predicted (Deng and Dodson 1994). Fibers with a log-normal distribution in length can be mathematically packed into regions, and the resulting structure and pore space characterized. Details of the sheet uniformity and pore space can be modeled with a computer. The complex nature of the fiber, such as its porosity and flexibility, still limits realistic predictions by computer models.

3. Uniformity of Porosity

During the formation of paper, water is drained through a filter media from a suspension of fibers and fillers to generate a mat. Modern machines promote drainage in two directions, top and bottom, in an attempt to make paper with the top and bottom surfaces the same. During drainage, the large fibers are easily captured on the machine filter media or "wire" while fine fibers and fillers move easily with the water phase. Common filtration theory indicates that as the filtercake increases in thickness, finer material can be caught. Therefore, in a one-dimensional drainage experiment there is often a gradient of fine material within a filtercake. The same phenomenon is observed with paper, in that fine material such as fiber fragments or filler particles are captured in some layers of the structure and depleted in other regions (Kajanto *et al.* 1998).

Another complexity with paper production is that the machine wire with the established filtercake will pass over a number of foils or drainage elements. Each element can actually force a small amount of water back into the filtercake, disrupt the filtercake, and remove the fine material deposited there. The net result is that paper is often depleted of fine material at its surfaces. If drainage is in one direction, there is often a large difference between the "wire" side and the "felt" side of the paper. Figure 2 depicts the drainage event and how fine material can deposit in different concentrations in different regions of the structure. The effect of filler location on paper properties is described by Allan *et al.* (1997). An analysis of variable fine particle retention within a filtercake is given in Tien *et al.* (1997).

A surface that is depleted of fines will tend to have a large pore size near the surface. This large pore size leads to low gloss of the sample, low image quality after printing, and a high ink demand to reach certain ink density targets during printing. As discussed, calendering can increase the gloss of paper and may influence the pore size of the surface layer, but too much calendering crushes the internal pores, causing poor opacity of the sheet.

The porosity of paper not only varies in the thickness direction due to the nonuniform distribution of fine material, but it can vary in the plane of the paper due to uniformity of the fiber mass distribution and the orientation of fibers. If the mass distribution of fibers or fines is not uniform in the plane of the paper, the "formation" of the paper is evaluated as "poor." A simple test of formation is to look through the paper with a light source on the other side. If there are regions of dark and light or it has a "cloudy" appearance, then the mass distribution of fiber is not uniform. The reason for the dark and light regions arises from the differing abilities of various regions to scatter light. This difference is a function of the porosity and the pore size of these regions. There is no standard method to characterize the variability of pore structure within the plane of the sheet.

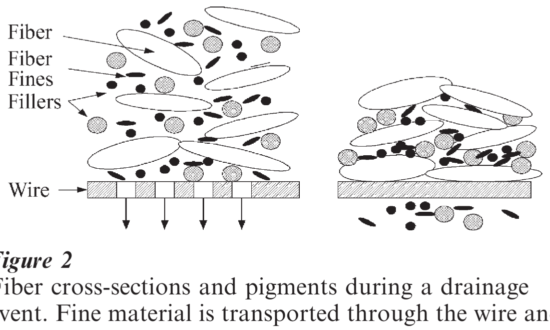


Figure 2
Fiber cross-sections and pigments during a drainage event. Fine material is transported through the wire and can be nonuniform in the thickness direction.

4. *Paper Coating*

In a number of grades of paper, a pigmented coating is applied to the surface of the paper to increase the optical and printing properties of the paper. The fiber small dimension is often around 30 μm, while coating pigment particles are usually less than 1 μm in diameter. The structure, roughness, and pore size of the coating layer is much finer than the base paper. Therefore, the coating increases the gloss, increases the opacity and brightness, and decreases the ink demand. The coating also will increase the possible resolution of the printed image. Figure 1 shows how the presence of a coating can increase the pore volume around a pore size of 0.1 μm. This pore volume is all at the surface of the paper, where a fine pore structure is needed for printing. This pore size is also good for light scattering, leading to an improved opacity of the product. Coating structure and pore size is detailed in Lepoutre (1988).

Aqueous paper coatings are often composed of kaolin or calcium carbonate pigments, latex binders, and other co-binders such as starch or protein. Depending on the quality of interest, coating weights of 5–30 $\mathrm{g\,m^{-2}}$ are normal. Coatings are often applied to the web at high speed with a pond or applicator roll, and the excess removed with a blade as shown in Fig. 3. The porosity of the base sheet influences this coating operation because capillary pressure and fluid pressure will drive the water phase and sometimes the pigment into the paper web. Highly porous base sheets lead to dewatering of the coating shortly after contact that can lead to operational problems at the blade and poor hold out of the coating layer on the paper surface. Therefore, the porosity of the base paper is important even in the coating operation.

The drainage of the coating suspension can create a situation similar to the drainage of paper fibers illustrated in Fig. 2: fine material can be redistributed within the thickness direction of the coating. This redistribution may generate a different pore structure

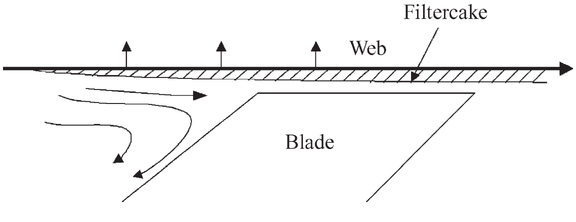


Figure 3
The blade metering process showing the loss of water from the coating into the paper web.

within the coating layer than that expected from coatings on plastic films. The movement of binder during the application or drying of coatings is called binder migration. This movement of binder is thought to have an impact on the final properties of the coating layer. The drying conditions associated with drying of the coating layer are known to influence the printing properties of the coating layer. A print defect called ‘‘back-trap mottle’’ is linked to the rapid drying of the coating layer. This defect must be generated by a nonuniform distribution of pores within the top surface of the coating layer.

5. Conclusions

The porosity of paper influences a wide range of properties important to the use of paper. Due to the complex nature of paper fibers and coating pigments, the pore space is difficult to describe in detail. A number of methods are available to characterize the porosity and the pore size distribution. The porosity of paper is determined by a number of process variables, but is a strong function of the fiber properties and any post-treatment of the paper web such as calendering. The porosity of paper is not uniform, but often varies both in the thickness direction and in the plane of the paper. Paper and paperboard are often coated with a pigmented coating to modify the surface pore structure. The porosity of the base web influences the coating operation, and the coating pore structure has an important impact on the printing process.

New techniques to measure and characterize paper porosity are being developed. Some techniques are able to measure the porosity of paper on-line. The use of computers to characterize paper structure has important implications for our ability to understand the complex pore structure in papers and paper coatings.

Bibliography

Allan G, Carroll J, Gaw K, Joseph A, Paramaporn D M, Pichitlamken J 1997 Effect of filler location on the drainage, pressing, and drying of pulp and paper. *TAPPI J* **80** (8), 175–9

Allem R, Uesaka T 1999 Characterization of paper microstructure: a new tool for assessing the effects of base sheet structure on paper properties. In: *Proc. TAPPI Advanced Coating Fundamentals Symp*. TAPPI Press, Atlanta, GA, pp. 111–20

Bureau W H 1976 Porosity from the inside out. *Graphic Arts Mon*. **48** (3), 86–8

Deng M, Dodson D T J 1994 *Paper: An Engineered Stochastic Structure*. TAPPI Press, Atlanta, GA

Kajanto I, Laamanen J, Kainulainen M 1998 Paper bulk and surface. In: Niskanen K (ed.) *Paper Physics*. Fapet Oy, Helsinki, pp. 89–115

Kershaw T N 1980 Sheet formation and drying. In: Casey J P (ed.) *Pulp and Paper, Chemistry and Chemical Technology*. Wiley, New York, pp. 945–1047

Lepoutre P 1988 *The Structure of Paper Coatings: An Update*. TAPPI Press, Atlanta, GA

Murakami K, Imamura R 1984 Porosity and gas permeability. In: Mark R E (ed.) *Handbook of Physical and Mechanical Testing of Paper and Paperboard*. Dekker, New York, Vol. 2, pp. 57–102

Niskanen K, Kajanto I, Pakarinen P 1998 Paper structure. In: Niskanen K (ed.) *Paper Physics*. Fapet Oy, Helsinki, pp. 14–53

Rance H F 1980 Consolidation of the paper web. In: Rance H F (ed.) *Handbook of Paper Science 1. The Raw Materials and Processing of Papermaking*. Elsevier, New York, pp. 209–88

Sayegh N N, Gonzalez T O 1995 Compressibility of fibre mats during drainage. *J. Pulp Paper Sci*. **21**, J255–61

Tien C, Bai R B, Ramarao B V 1997 Analysis of cake growth in cake filtration: effect of fine particle retention. *AIChE J* **43**, 33–44

D. W. Bousfield
University of Maine, Orono, Maine, USA

Paper: Structure

The mechanical behavior of a fiber network, such as a paper sheet composed of pulped wood fibers, depends on the fiber and fiber-to-fiber bond properties and the geometrical structure of the bonded fibrous network. The principal objective of a mathematical model is to predict the mechanical properties of the fiber network in terms of the properties of the fibers, bonds, and structure of the fiber network. In paper, the fibers are often collapsed, ribbon-like structures, which are bonded together primarily by hydrogen bonds formed when the sheet is pressed and subsequently dried.

The first attempt to develop a mathematical model for predicting the mechanical behavior of a fiber network such as paper was made by Cox (1952), who demonstrated how fiber orientation distribution influences the elastic behavior of the system. The analysis predicted all the elastic constants of a fiber mat in terms of the fiber orientation distribution expressed as a Fourier series.

Cox assumed that the fibers experience an axial strain corresponding to the sheet strain acting in the direction of the axis of the fiber. Cox’s original theory has been subjected to numerous refinements and extensions which take into account the structure of the fiber network, the modeling of the fibers, the effects of drying-induced strains, the modeling of bond deformation, the nonlinear constitutive effects of the deformation response of the fibers and bonds, and the effect of fiber buckling for fibers that experience compressive strains.

The Cox model was modified by Qi (1997) to account for transverse normal and shearing components of strain in the fiber by assuming that the fiber takes on the same strains as those of the sheet. This assumption is equivalent to that used in laminate theory; therefore, if the fibers are assumed to lie in the plane of the sheet, the Qi model gives the same result as that predicted by laminate theory.

Early work regarding the development of mathematical models can be found in Algar (1966), Dodson (1973), Kallmes (1972a, b), and Van den Akker (1970, 1972). General aspects of paper physics can be found in books by Niskanen (1998) and by Deng and Dodson (1994). A review of research involving the modeling of the mechanical behavior of paper was given by Ramasubramanian and Wang (1999).

In this article, the basic assumptions and results of current mathematical models for predicting the in-plane mechanical behavior of paper are presented. The approach that is presented is based on the mesomechanics of the fibers that comprise a fiber network following the work of Perkins (1983, 1990).

1. Basic Elements of the Mathematical Model

The choice of the mesoelement that is used in the construction of a mathematical model depends on the density of fiber sheet. In the case of low to medium density paper (to be defined subsequently), a typical mesoelement can be taken as an individual fiber along with the portions of crossing fibers that cross it and bond it to the fiber network. In order to develop a set of simple, closed-form expressions for the elastic moduli of the paper sheet, it is necessary to assume that the fibers are straight and that they lie in the plane of the paper. As discussed in the sequel, these assumptions can be relaxed, however, it is then necessary to perform the prediction of sheet properties numerically. Restricting attention to the low-density system comprised of straight fibers, the typical mesoelement of length, λ, and orientation, θ, is shown in Fig. 1. The mesoelements are coupled to the network by means of the crossing fibers. The strains in the sheet are presumed to be transmitted to the mesoelements by bending and shearing deformation of the crossing fibers and by shearing deformation of the fiber–fiber bonds. Thus the axial strain in the mesoelement fiber is not uniform but varies from the ends, where it is zero, to the middle, where its value is at its maximum. If the mesoelement is long enough and if the coupling is weak, the mesoelement strain is less than that associated with the sheet.

The model is further illustrated by Fig. 2, which shows a portion of a fiber that is coupled by two crossing fibers to the remainder of the network. The boundary between the element and the network is depicted by a dashed line. This boundary is assumed to be located a distance, $l/2$, from the centerline of the primary fiber, where l represents the center–center distance between bonds, l_b is the bond length along the fiber, and w_f and t_f are the width and thickness of the fiber, respectively. If l is small in comparison with w_f, as would be expected in moderately dense paper, the coupling is primarily attributable to the shearing deformations of the fiber–fiber bonds. In a very low-density system such as tissue paper, the bending and shearing deformation of the crossing fibers may be substantial.

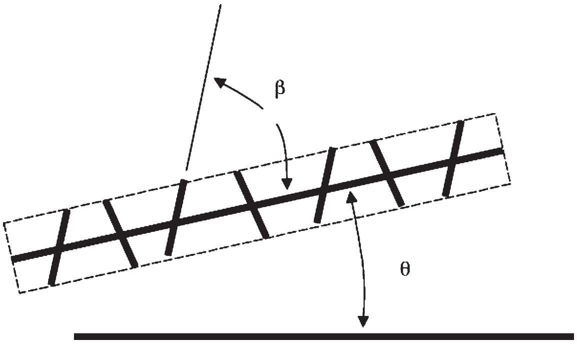


Figure 1
Typical fiber of length, λ_f, and orientation, θ, showing the angle of crossing fibers, β.

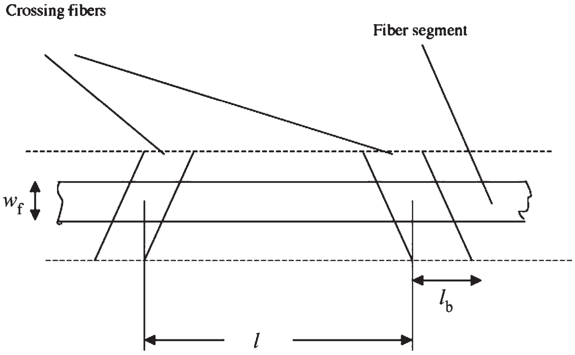


Figure 2
Portion of a fiber segment along a mesoelement showing the bond centroid distance, l, the bond length, l_b, and the fiber width, w_f. The mesoelement boundary applicable to low-density networks is shown as the dashed line.

On the assumption that the fiber is linearly elastic, the stress, σ_f, at any point along the fiber is related to the fiber strain, ε_f, at that point by

$$\sigma_f = E_{af}\varepsilon_f \quad (1)$$

where E_{af} is the effective axial modulus of the fiber. The effective axial modulus depends not only on the inherent properties of the fiber but also on how the fiber is bonded to other fibers in the network. Thus, there is a stiffening effect due to the reinforcement of other bonded fibers. E_{af} can be considered as an inherent property of the paper system, or its value can be estimated in terms of an idealized paper network.

For example, if the fibers are idealized (Figs. 1 and 2) as thin rectangular strips (e.g., perfectly collapsed thin-walled springwood fibers), and if the density of the network is low enough so that on average the fiber surface is only half (or less) in contact with other fibers in the network, then it can be shown that

$$E_{af} = \frac{(l/l_b)(E_{fL}+E_{fT})}{C_B\left(\dfrac{E_{fL}+E_{fT}}{E_{fL}}\right)\left(\dfrac{l-l_b}{l_b}\right)} \tag{2}$$

where

$$C_B = 1 + \frac{E_{fT}}{E_{fL}}\frac{\tanh B}{B} \tag{3}$$

and

$$B = \frac{l_b}{2}\sqrt{\left(\frac{k_b}{t_f}\left(\frac{E_{fT}+E_{fL}}{E_{fT}E_{fL}}\right)\right)} \tag{4}$$

Here E_{fl} and E_{fT} represent the elastic moduli in the axial and transverse directions of the fiber, respectively; t_f is the thickness of the collapsed fiber; and l and l_b are as previously defined.

The factor C_B determines the amount of reinforcement associated with fiber bonding. The constant k_b in the factor B (Eqn. (4)) determines the stiffness of the fiber–fiber bond. It is expected that these factors are related to the degree of fiber refining which roughens the fiber surface and causes damage or separation of the S1 layer of the cell wall. As the bond stiffness becomes very small, C_B tends to a value of 2 and the reinforcing effect vanishes. At the other extreme, the bond becomes infinitely stiff and the factor C_b tends to unity.

The bond stiffness can be measured directly as shown by Thorpe *et al.* (1976) where individual fibers are bonded to a fiber shive which is supported and the fiber is loaded by an externally applied axial load. The value of k_b measured in this way is on the order of $10^{13}\,\mathrm{Nm}^{-3}$.

The modeling of the elastic properties of the fiber depends on a number of factors related to whether or not the fibers are thin-walled and collapsed in the paper sheet and whether or not it is considered necessary to include the effects of the S_1 and S_3 layers of the cell wall. Refer to the work of Mark and Gillis (1983) and Salmen and de Ruvo (1985) for detailed information regarding the modeling of the fiber. Note that the transverse fiber modulus, E_{fT}, should be computed based on the crossing angle that the crossing fibers make relative to the primary fiber as discussed in the following.

The strain in the fiber, ε_f, is determined by the sheet strain and the degree of coupling between each fiber and the other fibers in the sheet. As shown in Fig. 1, a fiber of length λ_f is coupled to the rest of the network through the crossing fibers. The equilibrium equation for fiber mesoelement based on the equilibrium of a typical element along the fiber length is

$$\frac{d}{d\xi}\left(E_{af}\frac{d\varepsilon_f}{d\xi}\right) - k_f\varepsilon_f = -k_f\varepsilon_s \tag{5}$$

where

$$\varepsilon_s = \varepsilon_x\cos^2\theta + 2\varepsilon_{xy}\sin\theta\cos\theta \tag{6}$$

Equation (6) represents the normal component of strain in the sheet in the direction θ corresponding to the orientation of the fiber. The factor k_f is the effective stiffness of the coupling between the fiber of orientation θ and the other crossing fibers that couple the fiber to the network. The stiffness k_f includes the flexibility of the fibers in bending (which may be discernible for low-density sheets) and the flexibility of the fiber–fiber bond. If the idealized system of rectangular cross section straight fibers is considered,

$$k_f = \frac{2}{l\left[\dfrac{(l-l_b)^3}{12E_{fL}I_f}\left(1+\dfrac{12fE_{fL}I_f}{G_fA_f(l-l_b)^2}\right)+\dfrac{2}{A_bk_b}\right]} \tag{7}$$

where I_f represents the moment of inertia associated with the fiber cross-section, G_f represents the cell wall shear modulus, f is a factor that depends upon the fiber cross-sectional shape (it is 6/5 if the fiber is rectangular), A_b is the area of the fiber–fiber bond, and k_b is the bond stiffness.

Equations (2)–(4) and (7) can be used to estimate values of E_{af} and k_f in an idealized case. These quantities can also be considered as fundamental parameters of the system which are to be determined by mechanical property testing. In the latter case, direct measurement of paper mechanical response is used with the theoretical property relations to calculate "experimental" values of E_{af} and k_f.

It should be noted that the values of the effective axial modulus, E_{af}, and the coupling stiffness, k_f, are dependent on the orientation angle, θ, inasmuch as the network lengths, l and l_b, and the bond area, A_b, are dependent on orientation angle. In order to permit the subsequent closed-form expressions for the moduli of the sheet it is necessary to take the network lengths as their average values for the sheet. These values are dependent on the orientation distribution of the fibers as shown below.

For a prescribed sheet strain given by ε_x, ε_y, ε_{xy}, the normal component of sheet strain in the direction of the fiber, ε_s, is known. Equation (5) can be solved by assuming an arbitrary value of ε_s. The factors A_f, E_{af}, and k_f should properly be considered as varying in some random fashion. In this case, the solution to Eqn. (5) is very difficult. An alternative procedure is to select average values for A_f, E_{af}, and k_f and to determine an average fiber response. When this procedure is followed, the solution of Eqn. (5) is easily obtained as

$$\varepsilon_{\mathrm{f}} = \varepsilon_{\mathrm{s}}\left(1 - \frac{\cosh a\xi}{\cosh aL}\right) \tag{8}$$

$$a = \left(\frac{k_{\mathrm{f}}}{A_{\mathrm{f}}E_{\mathrm{af}}}\right)^{0.5} \tag{9}$$

L is half the fiber length and A_{f} is the cross-section area of the fiber.

The strain energy of a typical segment element is calculated as

$$W_{\mathrm{e}} = \tfrac{1}{2}E_{\mathrm{af}}A_{\mathrm{f}}\lambda\varepsilon_{\mathrm{s}}^{2}\eta_{\mathrm{L}} \tag{10}$$

where

$$\eta_{\mathrm{L}} = 1 - \frac{\tanh aL}{aL} \tag{11}$$

and can be identified as a coupling efficiency. For strong coupling and long length, η_{L} approaches unity. At the other extreme of very short length or very energy of the system can be written

$$W = \int_{0}^{x}\int_{0}^{\pi} D_{0}f_{\theta\lambda}^{*}W_{\mathrm{e}}\mathrm{d}\theta\mathrm{d}\lambda \tag{12}$$

where D_0 represents the number of segment elements per unit area and $f_{\theta\lambda}^{*}W_{e}d\theta d\lambda$ represents the probability of finding a segment having a length in the interval λ, $\lambda+\mathrm{d}\lambda$ and orientation in the interval θ, $\theta+\mathrm{d}\theta$.

Suppose that $f_{\theta\lambda}^{*}$ can be expressed in the form

$$\begin{aligned} f_{\theta}^{*} &= [f_{\theta}(\theta,\ a_1,\ a_2,\ a_3,\ \ldots,\ a_m)] \\ &\quad\times [f_{\lambda}(\lambda_s,\ b_1,\ b_2,\ b_3,\ \ldots,\ b_m)] \end{aligned} \tag{13}$$

Here, f_θ describes the fiber orientation distribution as a function of θ depending on the m parameters, a_1, a_2, …, a_m. Likewise, the distribution of fiber lengths, λ, depends on n parameters b_1, b_2, …, b_n. In general, the parameters a_1, a_2, …, a_m are assumed to be functions of λ, while the parameters b_1, b_2, …, b_n are functions of θ.

The evidence available suggests that coupling between the two distributions is weak and therefore as an approximation, valid for at least certain papers, it can be assumed that the length and orientation distributions are independent. For this assumption, the parameters a_1, a_2, …, a_m and b_1, b_2, …, b_n are constants, and the strain energy per unit sheet area can be written in the form

$$W = \frac{1}{2}\frac{\omega_{\mathrm{s}}}{\rho_{\mathrm{af}}}\int_{0}^{\pi}\phi_{\lambda}E_{\mathrm{af}}\varepsilon_{s}^{2}f_{\theta}\mathrm{d}\theta \tag{14}$$

where ω_{s} represents the basis weight, E_{af} represents the apparent fiber density and

$$\phi_{\lambda} = \int_{0}^{\infty}\eta_{\mathrm{L}}f_{\lambda}\mathrm{d}\lambda \tag{15}$$

The length of parameter ϕ_λ and the apparent fiber modulus, E_{af}, may depend upon the orientation direction, θ, as a result of the stresses imposed on the fibers during drying of the sheet. Analytically, it can be assumed that the apparent fiber modulus, E_{af}, depends upon the shrinkage and restraint conditions present during drying in the following way:

$$E_{\mathrm{af}} = E_{\mathrm{af}_0}(1 + H\varepsilon_{\mathrm{ND}}) \tag{16}$$

where E_{af_0}, a constant, represents the apparent fiber modulus in the absence of drying restraint, H is a constant that predicts the magnitude of stiffening, and

$$\varepsilon_{\mathrm{ND}} = (\alpha_x M + \varepsilon_{x\mathrm{D}})\cos^2\theta + (\alpha_y M + \varepsilon_{y\mathrm{D}})\sin^2\theta \tag{17}$$

Here $\alpha_x M$ and $\alpha_y M$ represent the sheet shrinkage strains during unrestrained drying in the x and y directions, and $\varepsilon_{x\mathrm{D}}$ and $\varepsilon_{y\mathrm{D}}$ represent the sheet strains applied or allowed during drying. Thus, for unrestrained shrinkage conditions, $\varepsilon_{x\mathrm{D}} = -\alpha_x M$ and $\varepsilon_{y\mathrm{D}} = -\alpha_y M$. It is evident that $\varepsilon_{\mathrm{ND}}$ is related to the magnitude of restraint that a fiber of orientation θ would be subjected to during the drying of the sheet. In the following analysis, it is further assumed that $\varepsilon_{\mathrm{ND}}$ is positive, that is, the fibers are loaded in tension as a result of any drying restraint.

Experimental results of the effects of drying restraint were performed by Setterholm and Chilson (1965). Data taken from their work (cf. Perkins and Mark (1981)) indicates that the parameter H has values that lie in the range of 10 to 30. Experimental evidence of the effect of drying restraint on the properties of the fibers in the network was shown by Wuu *et al.* (1991).

With the help of Eqn. (16), integration of Eqn. (14) can be easily carried out, if the length factor, ϕ_λ, is independent of θ. In fact, ϕ_λ does depend on θ, because the coupling parameter a that appears in η_{L} depends on E_{af}. If one assumes that the influence of variation in E_{af} through θ in ϕ_λ has an insignificant effect on the prediction of the elastic moduli of the sheet, there is no serious analytical error committed in assuming the ϕ_λ is independent of θ.

The stress–strain relations for the sheet can be obtained from the strain energy function W after carrying out the integration through the relations

$$\tau_x = \frac{\partial W}{\partial \varepsilon_x}\tau_y = \frac{\partial W}{\partial \varepsilon_y}\tau_{xy} = \frac{1}{2}\frac{\partial W}{\partial\,\varepsilon_{xy}} \tag{18}$$

where τ_x, τ_y, τ_{xy} represent the force per unit edge length of the paper sheet.

To express conveniently the resulting elastic moduli it is desirable to express the orientation distribution $f_\theta(\theta)$ in terms of its Fourier series expansion. When the x direction is one of the axes of elastic symmetry, for example the machine direction, and

$\theta = 0^\circ$ corresponds to this direction,

$$f_\theta(\theta) = \frac{1}{\pi}[1 + a_1 cos2\theta + a_2 cos4\theta + a_3 cos6\theta + \cdots + a_n cos2n\theta + \cdots] \quad (19)$$

After substitution of Eqns. (19) and (16) in Eqn. (14) and subsequently using Eqn. (18), the Young's moduli E_x^*, E_y^*, the shear modulus G_{xy}^*, and the Poisson ratios $\upsilon_{xy,}\upsilon_{yx}$ corresponding to plain stress loading of the sheet are found to be

$$E_x^* = \frac{1}{16}\frac{\omega_\mathrm{s}}{\rho_\mathrm{af}}\phi_\lambda E_{\mathrm{af}_0}\Big[(1 + \langle e \rangle)(6 + 4a_1 + a_2). + \frac{e_\Delta}{2}(8 + 7a_1 + 4a_2 + a_3)\Big][1 - v_{xy}v_{xy}] \quad (20a)$$

$$E_y^* = \frac{1}{16}\frac{\omega_\mathrm{s}}{\rho_\mathrm{af}}\phi_\lambda E_{af_0}\Big[(1 + \langle e \rangle)(6 - 4a_1 + a_2). + \frac{e_\Delta}{2}(8 - 7a_1 + 4a_2 - a_3)\Big][1 - v_{xy}v_{xy}] \quad (20b)$$

$$v_{xy} = \frac{(1 + \langle e \rangle)(2 - a_2) + \frac{e_\Delta}{4}(a_1 - a_3)}{(1 + \langle e \rangle)(6 - 4a_1 + a_2) - \frac{e_\Delta}{2}(8 - 7a_1 + 4a_2 - a_3)} \quad (20c)$$

$$v_{yx} = \frac{(1 + \langle e \rangle)(2 - a_2) + \frac{e_\Delta}{4}(a_1 - a_3)}{(1 + \langle e \rangle)(6 + 4a_1 + a_2) + \frac{e_\Delta}{2}(8 + 7a_1 + 4a_2 + a_3)} \quad (20d)$$

$$G_{xy}^* = \frac{1}{16}\frac{\omega_s}{\rho_\mathrm{af}}\phi\lambda E_{\mathrm{af}_0} \times [(1 + \langle e \rangle)(2 - a_2) + \frac{e_\Delta}{2}(a_1 - a_3)] \quad (20e)$$

where

$$\langle e \rangle = \tfrac{1}{2}H[(\alpha_x M + \varepsilon_{x\mathrm{D}}) + (\alpha_y M + \varepsilon_{y\mathrm{D}})] \quad (21a)$$

$$e_\Delta = \tfrac{1}{2}H[(\alpha_x M + \varepsilon_{x\mathrm{D}}) - (\alpha_y M + \varepsilon_{yD})] \quad (21b)$$

An inspection of Eqns. (20a–e) indicates that the elastic constants depend only on the first three parameters, a_1, a_2, a_3, that occur in the Fourier series expansion for f_θ. Furthermore, if the influence of drying restraint is not incorporated or if the sheet is dried under conditions of no restraint, the elastic moduli depend only on the parameters a_1 and a_2. For this condition, Eqns. (20a–e) are essentially the same as those given by Cox (1952) when $\theta = 0^\circ$ corresponds to a direction of elastic symmetry in the sheet. (Note, however, that the effects of fiber–fiber bonding and the effect of finite fiber length are taken into account.)

The number of coefficients of the Fourier series expansion for f_θ that are necessary for predicting the elastic properties, therefore, depends on the functional relation between the apparent fiber modulus, E_af, and orientation, which in turn depends on the procedures and conditions of drying. Incorporation of any other phenomena affecting the angular depencence of the strain energy stored in the sheet may change the number of Fourier coefficients appearing in the elastic constants.

It is theoretically possible to describe any fiber orientation distribution in terms of the parameters a_1, a_2, a_3, … of the Fourier series expansion. However, a single-parameter function would simplify the expressions for the elastic moduli. The reduction to one parameter can be accomplished very simply by truncating the Fourier series and retaining the single parameter a_1. However, reduction to a coefficient in the Fourier series expansion can have some significant effects. For example, for freely dried paper, the shear modulus should decrease with increased anisotropy, but this reduction effect will be predicted only if the coefficient a_2 is retained in the expansion of Eqn. (19). A single-parameter function would predict the shear modulus to be independent of the degree of anisotropy of the paper.

An alternative approach is to employ a distribution function that has only one shape parameter. For a general discussion of fiber orientation distribution refer to Perkins and Mark (1981). The two most suitable single-parameter distributions are the elliptical and the von Mises distributions. The distributions and their corresponding values of a_1, a_2, a_3 are given in Table 1.

2. *Application of Theory to Prediction of Elastic Moduli of a Paper Sheet*

Equations (20a–e) can be used to estimate the effects of changes in structural parameters on the elastic properties of paper, or to estimate from experimental measurements the parameters of the system. As an example, if the values of E_x and E_y can be obtained from measurements of the Young's moduli in the machine and cross-machine directions, the remaining in-plane elastic moduli can be estimated.

To use Eqns. (20a–e) in this example, it is necessary to investigate the influence of drying restraint. The stiffening parameter H of Eqn. (21) has been estimated by experiment (Setterholm and Chilson 1965) to lie in the range 10–30. From a knowledge of the drying restraint conditions and sheet shrinkage behavior, the values of $\langle e \rangle$ and $\langle e_\Delta \rangle$ can be calculated. The ratio of the experimental Young's moduli for the machine and cross-machine directions can be equated to the ratio E_x^*/E_y^* from Eqns. (20a,b). The

Table 1
Parameters for the elliptical and von Mises fiber orientation distributions.

	Parameter		
	a_1	a_2	a_3
Elliptical $f_\theta = \frac{\zeta}{\pi}\left(\frac{1}{\cos^2\theta+\zeta\sin^2\theta}\right)$	$2\left(\frac{\zeta-1}{\zeta+1}\right)$	$2\left(\frac{\zeta-1}{\zeta+1}\right)^2$	$2\left(\frac{\zeta-1}{\zeta+1}\right)^3$
von Mises[a] $f_\theta = \frac{1}{\pi I_0(\kappa)}e^{\kappa cos2\theta}$	$2\frac{I_1(\kappa)}{I_0(\kappa)}$	$2\frac{I_2(\kappa)}{I_0(\kappa)}$	$2\frac{I_3(\kappa)}{I_0(\kappa)}$

[a] $I_n(\kappa)$ = modified Bessel function of the first kind and order n.

elliptical or von Mises distribution can be written in a Fourier series form (Table 1). The first three coefficients of the expansion can then be taken as a_1, a_2, a_3, respectively. The inversion of this relation would yield a value of the concentration parameters for orientation ζ or κ can then be used to obtain a_1, a_2, a_3 and subsequently the values of v_{xy}, v_{yx}, and G_{xy}. The measurement of two Young's moduli then permits all the in-plane elastic moduli for the sheet to be estimated. The calculation is best performed for several different assumptions regarding the variable in Eqn. (21), to ascertain the reliability of the assumptions.

The choice of fiber orientation distribution function can have significant effects on predictions obtained from Eqns. (20a–e). For a given fiber orientation, predictions of v_{xy}, v_{yx}, and G_{xy} from experimental measurements of E_x and E_y may be different for the elliptical distribution from those for the von Mises distribution. While the elliptical distribution or the single-parameter cosine distribution is the easiest to apply, the von Mises distribution is considered to provide more accurate estimates (Perkins and Mark 1981).

In the case of fiber segment length distributions, the Erlang distribution has been found to be quite satisfactory. This, the function f_{λ_s} can be taken as

$$f_{\lambda_s} = \frac{(\lambda_s/b)^{c-1}\exp(-\lambda_s/b)}{b[(c-1)!]} \tag{22}$$

Here, b and c are determined by fitting the distribution to the experimental data. The product bc is equal to the mean value of λ_s. The standard deviation for the distribution is $b\sqrt{c}$. For further information, refer to Perkins and Mark (1981).

The distance between fiber crossings of fiber-to-fiber bonds along a given fiber, $l(\theta)$, is an important property of the fiber network. Assuming that the fibers are all straight, the bond centroid distance depends on the direction of the fiber and the overall orientation distribution of the fibers in the network. The bond centroid distance is determined by the number of fibers that cross a given fiber. This problem was studied by Kallmes and Corte (1960) and by Komori and Makishima (1977) who showed that

$$n(\theta) = \frac{2D_\mathrm{f}N\lambda^2}{V}J(\theta) \tag{23}$$

where

$$J(\theta) = \int_0^\pi f(\theta)|\sin(\theta-\alpha)|\mathrm{d}\alpha \tag{24}$$

Here, D_f denotes the fiber diameter, or in the case of a flattened fiber, the z-direction thickness of the fiber. N represents the number of fibers in a volume V. The function $J(\theta)$ represents the average value of the sine of the crossing angle, $\theta-\alpha$, for a fiber of orientation θ. The function $J(\theta)$ can be expressed in series form as shown by Chang (1983)

$$J(\theta) = \frac{2}{\pi}\left[1 - \sum_{n=1}^{\infty}\frac{a_n}{(2n-1)(2n+1)}cos(2n\theta)\right] \tag{25}$$

The rationale for Eqn. (23) follows from the observation that $2D_\mathrm{f}\lambda(\lambda\sin\beta)$ represents the volume occupied by the two crossing fibers times the projected length of the crossing fiber. Thus, the average number of contacts is the number of fibers, N, in a volume, V, times the ratio of the volume associated with one crossing to the total volume of fiber material.

The average distance between bond centroids, $l(\theta)$, along a fiber of orientation θ is taken as the reciprocal of the average number of crossings per unit length. Therefore,

$$l(\theta) = \frac{V}{2D_\mathrm{f}N\lambda J(\theta)} \tag{26}$$

The average number of crossings considered over the set of all fibers is

$$\bar{n} = \frac{2D_f N\lambda^2}{V}I \tag{27}$$

where

$$I = \int_0^\pi f(\alpha)J(\alpha)\mathrm{d}\alpha \tag{28}$$

or in series form as shown by Chang (1983)

$$I = \frac{2}{\pi}\left[1 - \sum_{n=1}^{\infty}\frac{a_n}{(2n-1)(2n+1)}\right] \tag{29}$$

The overall average distance between bond centroids for the network is

$$l = \frac{V}{2D_f N\lambda I} \tag{30}$$

The quantity $N\lambda$ represents the total length of fiber in a volume, V. If $M = \rho_f A_f N\lambda$ represents the mass in volume V, then the expressions for bond centroid distance can be written,

$$l(\theta) = \frac{\rho_f}{2\rho_s}\frac{A_f}{D_f}\frac{1}{J(\theta)} \tag{31}$$

$$l = \frac{\rho_f}{2\rho_s}\frac{A_f}{D_f}\frac{1}{I} \tag{32}$$

When the fibers are assumed to be completely flattened, then D_f is the thickness t_f and $A_f = w_f t_f$ and the bond centroid distances can be written

$$l(\theta) = w_f\frac{\rho_f}{2\rho_s}\frac{1}{J(\theta)} \tag{33}$$

$$l = w_f\frac{\rho_f}{2\rho_s}\frac{1}{I} \tag{34}$$

The fiber crossing angle can be obtained from the definition of $J(\theta)$ since this quantity represents the average value of the sine of the angle between two crossing fibers for a fiber of given orientation angle θ relative to the reference direction. Therefore, the average crossing angle β is given by

$$\beta = \arcsin(J(\theta)) \tag{35}$$

The bond area is calculated from the crossing angle. Therefore,

$$A_b(\theta) = \frac{w_f^2}{J(\theta)} \tag{36}$$

and the average bond area for fibers of all orientations is

$$A_b = \frac{w_f^2}{I} \tag{37}$$

These network properties are to be used in the calculations for the fiber coupling parameter k_f and the effective fiber modulus E_{af} in Eqns. (2) and (7), respectively.

When the relations provided in Eqn. (20) are to be used to predict the sheet elastic moduli, the values of l, and A_b are to be taken from Eqns. (34) and (37) because these relations relate the average properties of the sheet over all directions. The portion of the fiber that is bonded, l_b, is defined using Eqns. (36) and (37) as the apparent length necessary in order that the product $l_b w_f$ is equal to the bond area. On the other hand, the methods leading to Eqn. (20) can be performed numerically with the angle dependence of bond centroid distance, bond length, and bond area taken into account using Eqns. (33) and (36) in the expressions for E_{af} and k_f. The numerical approach makes it possible to avoid the simplifying assumption of independence of angular dependent and fiber length effects that are discussed above.

Using the relations above, it can be shown that when the sheet density, ρ_s, is equal to one-half the fiber density, ρ_f, the fibers will be in contact with adjacent fibers on one side of the fiber or the other and the unbonded length goes to zero. Given that the fiber density is $1500\,\mathrm{kg\,m^{-3}}$, the network relations above predict that fibers will have no unbonded or free length when the sheet density is above $750\,\mathrm{kg\,m^{-3}}$. For paper sheets characterized by a density in excess of this figure, the mesoelement model depicted by Fig. 1 should be modified so that there is no free fiber length. This would suggest that the mesoelement should be taken as the fiber itself. Therefore, for sheet density in excess of $750\,\mathrm{kg\,m^{-3}}$ the relations for the elastic moduli (Eqn. (20)) can still be used as long as the effective axial fiber modulus from Eqn. (2) is taken as the axial fiber modulus, E_{fL}, and the network coupling stiffness, k_f, is taken as

$$k_f = \frac{A_b k_b}{l} \tag{38}$$

When the procedure described above is used for predicting the elastic moduli of sheets with sheet density in excess of $750\,\mathrm{kg\,m^{-3}}$ it is clear that the prediction neglects the transverse and shearing strains in the fiber. The laminate approach was explored by Schulgasser and Page (1988) and by Page and Schulgasser (1989) in order to take into account the transverse and shearing contribution. They showed that the transverse properties of fibers are especially important as regards the prediction of the in-plane shear modulus and the Poisson ratios of the sheet. In the works of Lu *et al.* (1995), Lu and Carlsson (1996), and Lu *et al.* (1996) the authors develop what they refer to as a mosaic model that incorporates the in-plane axial and transverse elastic moduli of the fiber. This approach makes use of the finite element method to analyze a representative element that is characteristic of the fiber segments and bonds that comprise the paper network. A model for moderate to high density paper was also proposed by Uesaka (1990) where the transverse strain effects were taken into account by modifying the theory leading to Eqns. (20) to incorporate transverse fiber strain in a similar fashion to the modeling of the axial strain. This work takes into account the contribution of the fiber-to-fiber bonds, but neglects the fiber Poisson and shearing effects. An approximate model that accounts for transverse strain, Poisson and shearing effects, and the fiber–fiber bonding has been proposed by Perkins (1999).

3. Nonlinear Geometry and Constitutive Effects

The assumption of linear elastic behavior of the fiber and bond materials can be relaxed while using essentially the same approach as that outlined above. The mechanical response of paper when loaded uniaxially, as in the example of a simple tensile test, was studied by Ramasubramanian and Perkins (1988) who developed algebraic expressions for predicting the uniaxial stress–strain response. The expressions are sufficiently complex that their evaluation by computer is necessary. This work was extended to account for moderate to high density papers by Perkins *et al.* (1991). Subsequently, Sinha and Perkins (1995) adapted the mesoelement approach to provide a constitutive model for use in finite element analysis.

The basic approach outlined above can be modified to take into account a number of nonlinear effects if one does not require that closed form expressions be obtained for the prediction of mechanical response. In other words, the mesoelement approach can be extended by performing the calculations numerically rather than in algebraic form. The effects of fiber curl, fiber buckling, nonelastic constitutive models for the bond and fiber components, as well as certain aspects of failure at the mesoelement level can be studied when the mesoelement model is analyzed by the finite element method. The basic approach has been presented by Perkins (1999).

See also: Paper: Porosity

Bibliography

Algar W H 1966 In: Bolam F (ed.) *Consolidation of the Paper Web*. Technical Section of the British Paper and Board Makers' Association, Cambridge, Vol. 2, pp. 814–49

Chang J 1983 Syracuse University, Syracuse, NY

Deng M, Dodson C T J 1994 *Paper: An Engineered Stochastic Structure*. Tappi Press

Dodson C T J 1973 In: Bolam F (ed.) *The Fundamental Properties of Paper Related to its Uses. Technical Association of the British Paper and Board Makers' Association*. Cambridge, Vol. 1, pp. 202–26

Kallmes O 1972a In: Jayne B A (ed.) *Theory of Design of Wood and Fiber Composite Materials*. Syracuse University Press, Syracuse, NY, pp. 157–76

Kallmes O 1972b In: Jayne B A (ed.) *Theory of Design of Wood and Fiber Composite Materials*. Syracuse University Press, Syracuse, NY, pp. 177–96

Kallmes O, Corte H 1960 The structure of paper, I. *Tappi*. **43**, 737–52

Komori T, Makishima K 1977 *Textile Res. J*. **47**, 13–7

Lu W, Carlsson L A 1996 Micromodel of paper, II. *Tappi J*. **79**, 205–10

Lu W, Carlsson L A, Andersson Y 1995 Micromodel of paper. *Tappi J*. **78**, 155–64

Lu W, Carlsson L A, de Ruvo A 1996 Micromodel of paper, I. *Tappi J*. **79**, 197–203

Mark R E, Gillis P P 1983 In: Mark R E (ed.) *Handbook of Physical and Mechanical Testing of Paper and Paperboard*. Marcel Dekker, New York, Vol. 1, pp. 409–95

Niskanen K (ed.) 1998 *Paper Physics*. Fapet Oy, Hensinki, Finland

Page D H, Schulgasser K 1989 In: Perkins R W (ed.) *Mechanics of Cellulosic and Polymeric Materials*. ASME, New York, Vol. AMD-Vol. 99, MD-Vol. 13, pp. 35–9

Perkins R 1990 In: Caulfield D F, Passaretti J D, Sobczynski S F (eds.) *Materials Interactions Relevant to the Pulp, Paper, and Wood Industries*. Materials Research Society, San Francisco, CA, Vol. 197, pp. 99–118

Perkins R W 1983 In: Mark R E (ed.) *Handbook of Physical and Mechanical Testing of Paper and Paperboard*. Marcel Dekker, New York, Vol. 1, pp. 23–5

Perkins R W 1999 *ASME Summer Mechanics Conference-Symposium on Mechanics of Cellulosic Materials*, Vol. AMD-Vol. 231

Perkins R W (ed.) *American Society of Mechanical Engineers*, Blacksburg, VA, MD-Vol. 85

Perkins R W, Mark R E 1981 The role of fundamental research in papermaking. In: Brander J (ed.) *Trans. Symp*. Cambridge: September *1981,* Mechanical Engineering Publications Limited, London, Cambridge, UK, Vol. 1, pp. 479–525

Perkins R W, Sinha S, Mark R E 1991 In: Mark R E (ed.) *Int. Paper Physics Conf.*. TAPPI, Kona, Hawaii, Vol. 2, pp. 413–35

Qi D 1997 *Tappi J*. **80**, 165–71

Ramasubramanian M K, Perkins R W 1988 *J. Eng. Mater. Technol*. **110**, 117–23

Ramasubramanian M K, Wang Y Y 1999 In: Perkins R W (ed.) *Mechanics of Cellulosic Materials*. ASME, New York, pp. 31–42

Salmen L, de Ruvo A 1985 *Wood Fiber Sci*. **17**, 336–50

Schulgasser K, Page D H 1988 *Composites Sci. Technol*. **32**, 279–92

Setterholm V C, Chilson W A 1965 *Tappi J*. **48**, 634–40

Sinha S S, Perkins R W 1995 In: Perkins R W (ed.) *Mechanics of Cellulosic Materials*. ASME, New York, Vol. MD-Vol 60, AMD-Vol 209

Thorpe J, Mark R E, Eusufzai A R K, Perkins R W 1976 *Tappi J*. **59**, 96–100

Uesaka T 1990 In: Perkins R W (ed.) *Mechanics of Wood and Paper Materials*. ASME, New York, Vol. AMD-Vol. 112, MD-Vol. 23, pp. 29–35

Van den Akker J A 1970 *Tappi J*. **53**, 388–400

Van den Akker J A 1972 In: Jayne B A (ed.) *Theory of Design of Wood and Fiber Composite Materials*. Syracuse University Press, Syracuse, NY, pp. 197–218

Wuu F, Mark R E, Perkins R W 1991 In: Mark R E (ed.) *International Paper Physics Conference*. Tappi, Kona, Hawaii, Vol. 2, pp. 663–71

R. Perkins
Syracuse University, New York, USA

Pearlite

Pearlite forms as the result of the eutectoid reaction and consists of alternate lamellae of ferrite (α) and cementite (Fe_3C, θ) growing synchronously and cooperatively into the austenite (γ) phase. It is certainly the most recognizable, most esthetically pleasing, and the most commercially important eutectoid phase transformation. Because of its historical importance ferrous pearlite has become the prototype for nonferrous pearlites as well. Pearlite was first described by Sorby in 1886 and has since gone through several major changes in the description of its formation. Because pearlite involves the nucleation and growth of two new phases, which ultimately cooperate, this phase transformation has unique characteristics that distinguish its development. The well-established features of pearlite formation will be outlined in this article and details concerning the role of thermodynamics and the importance of crystallography in the development of a pearlite colony are presented. Emphasis will be on binary Fe–C alloys; however, Fe–C–X alloys will also be discussed.

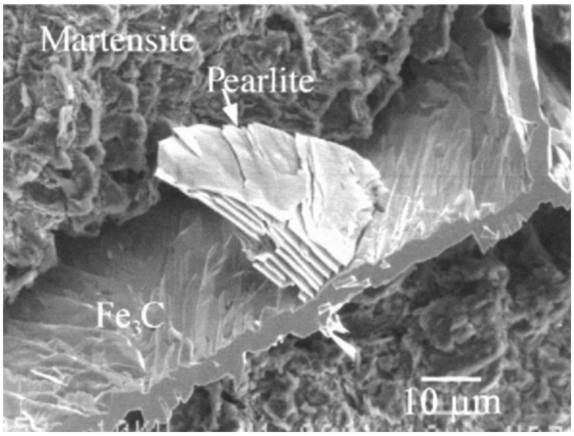


Figure 1
Scanning electron micrograph of a pearlite colony grown from proeutectoid cementite. All the ferrite and much of the matrix are etched away. The plates of cementite lamellae are attached to a film of grain boundary proeutectoid cementite film.

1. Morphology and Historic Development

Pearlite is generally observed to nucleate at austenite grain boundaries and grow into the parent phase as approximately spherically shaped grains. These grains are called nodules and are composed of a number of structural units called colonies in which all of the plates are parallel. Because of crystallographic constraints it is often observed that the pearlite nodules only grow into one of the adjoining grains. Figure 1 shows the parallel plates of a colony in a hypereutectoid steel where all the pearlitic ferrite is etched away. The radial nature of the colony is revealed, as well as the connection of the cementite lamella with the proeutectoid grain boundary cementite. When growing colonies intercept grain and twin boundaries, individual lamella either are halted or abruptly change growth direction upon impingement. At high reaction temperatures, "group nodules" of pearlite may form that consist of many colonies of pearlite. The overall morphology of a group nodule may be approximately described as spherical. At lower reaction temperatures, pearlite nodules, again consisting of many colonies, form as hemispheres or as sectors of spheres. This morphology is known as "grain boundary" pearlite or as "nodular fine pearlite."

In an Fe–C alloy of eutectoid composition, pearlite is the only product formed from the eutectoid temperature, 727 °C, down to about 600 °C. Small amounts of bainite appear at 600 °C. The proportion of bainite increases with decreasing temperature, and at 500 °C the amounts of pearlite and bainite are about equal. The pearlite reaction then disappears rapidly. In hypo-and hypereutectoid steels, pearlite forms at similar temperatures after the proeutectoid constituent has formed whose proportion diminishes with decreasing temperature. A hypoeutectoid steel containing as little as 0.35% carbon can be wholly converted to pearlite.

Since the work of Sorby, major contributions to the understanding of pearlite nucleation and growth have been made by Mehl and Hagel (1956) and by Hillert (1962). Each of these explanations was the primary and generally accepted description of pearlite transformation from the 1930s to the 1950s for the former and since the 1960s for the latter. In Mehl's repeated nucleation and growth mechanism of pearlite formation, an initial plate of cementite forms at an austenite grain boundary. Following this, it is thermodynamically and energetically (a low-energy interfacial structure between ferrite and cementite minimizes the nucleation barrier for each phase) favorable for a ferrite plate to nucleate on either side of the cementite plate. This is followed by cementite plate nucleation, and so on, resulting in sideways growth, i.e., increased volume of the pearlite colony perpendicular to the lamellae. Although still pedagogically attractive, Mehl's mechanism has not been verified by sectioning experiments. Hillert (1962), however, pointed out that the growth of pearlite is a gradual process with the lamellae branching and searching for an optimum or favorable orientation between them.

Once found, rapid cooperative growth ensues because of the highly mobile incoherent interface. On the question of branching vs. independent nucleation of precipitate lamellae, repeated sectioning experiments point conclusively toward the branching explanation (Hillert 1962, Mangan and Shiflet 1999). Hackney and Shiflet (1987a, b) added crystallographic considerations at the growth interface to explain pearlite edgewise growth behavior. On their observations, growth ledges migrate across the pearlite/austenite interphase boundary and are shared between ferrite and cementite. The lamellae of these two phases are thus forced to lengthen at the same rate. Because the pearlite growth front requires ledges, many pearlite phenomena, such as synchronous growth, the inability of individual lamellae to penetrate grain and twin boundaries unaffected, and the precipitation of interphase boundary carbides, can be simply explained.

2. Thermodynamic Considerations

Pearlite nucleation is best visualized on the basis of the Fe–Fe_3C phase diagram with extrapolated phase boundaries (Fig. 2). An alloy of hypoeutectoid composition cooled from the austenite phase field to temperature T_1 provides favorable conditions for ferrite nucleation. As a result of kinetic considerations, proeutectoid ferrite can still form at temperatures below that of the eutectoid. This is represented by the extrapolated phase boundaries below the eutectoid temperature. After grain boundary ferrite forms, the carbon concentration at its boundaries is that of the extrapolated Ae_3 curve, $x_\gamma^{\gamma\alpha}$. Substantial supersaturation is now present for cementite nucleation at the ferrite/austenite boundaries. When a cementite particle nucleates at this boundary, the carbon concentration at the austenite/cementite (θ) boundary, $x_\gamma^{\gamma\theta}$, is that of the extrapolated A_{cm} curve. Both the volume free energy and interfacial energy conditions are now excellent for the nucleation of another crystal of ferrite. This ferrite crystal nucleates with deference to the cementite crystal establishing one of the three orientation relationships.

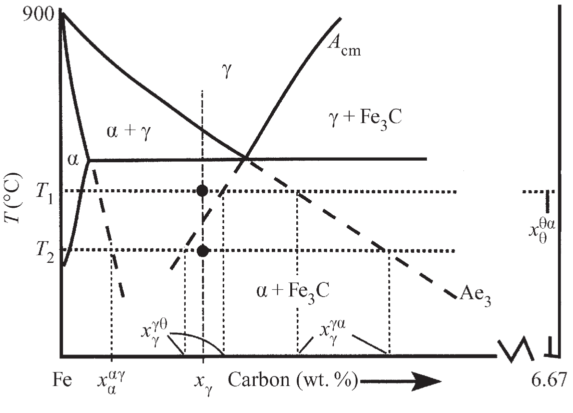


Figure 2
Portion of the eutectoid region of the Fe–C phase diagram with extrapolated Ae_3 and A_{cm} and $\alpha/(\alpha+\gamma)$ phase boundaries. Various compositions are indicated at two isotherms, T_1 and T_2.

At a lower temperature (T_2) the alloy lies between the extrapolated Ae_3 and A_{cm} and there is sufficient supersaturation with respect to both ferrite and cementite nucleation so that both can occur at the grain boundary. Hence, it may be possible to form an entirely pearlitic microstructure in an off-eutectoid alloy if it is quenched to temperatures below the extrapolated A_{cm}. Similar arguments can be made for hypereutectoid alloys where at small undercoolings grain boundary cementite appears first.

3. Nucleation and Crystallography

Because pearlite is the result of a solid–solid phase transformation, nucleation cannot be separated from crystallography. In Fe–C pearlite the two lamellar components are composed of b.c.c. ferrite and complex orthorhombic cementite (Fe_3C, here the convention that the lattice parameters $a<b<c$ is adopted). The association of these two structures has consistent and reproducible effects beyond forming a well-defined lamellar phase. Pearlite can nucleate at three different sites, namely in association with proeutectoid ferrite, at austenite boundaries where no precipitation has already occurred, and at proeutectoid cementite interphase boundaries. Although the importance and number of orientation relationships reported has fluctuated over the years, three have been consistently reported in the literature. They are the Bagaryatsky (1950), Isaichev (1947), and Pitsch (1962)–Petch (1953) orientation relationships, as expressed in Table 1. Analysis reveals that the Bagaryatsky and Isaichev relationships are very similar with only a few degrees of rotation around the b-axis of cementite. As shown in Table 1, however, each orientation relationship yields unique interlamellar habit planes. Hence, their distinction should be preserved (Zhou and Shiflet 1992).

The relative frequency with which the three orientation relationships occur in pearlite depends upon alloy composition and thermal history. The general trends for all three have not yet been established. Unfortunately, most reports in the literature are incorrect because of reliance on essentially two-dimensional sections (electron transparent foil) of a pearlite colony to interpret its three-dimensional structure. These studies can misrepresent the actual situation. Mangan and Shiflet (1999) have shown with a combination of electron backscattered diffraction using scanning electron microscopy and repeated sectioning of entire pearlite colonies that the Pitsch–Petch relationship develops when a colony is initiated by ferrite nucleating at a cementite/austenite interphase boundary. This occurs in both hypo- and hypereutectoid alloys. The nucleation conditions for establishing

Table 1
Orientation relationships and interlamellar habit planes of ferrous pearlite.

Relationship	Orientation relationship	Habit plane
Bagaryatsky (1950)	$[100]_\theta \| [\bar{1}10]_\alpha$ $[010]_\theta \| [111]_\alpha$ $(001)_\theta \| [11\bar{2}]_\alpha$	$(001)_\theta \| (112)_\alpha$
Isaichev (1947)	$[010]_\theta \| [111]_\alpha$ $(101)_\alpha \| (11\bar{2})_\alpha$	$(101)_\theta \| [11\bar{2}]_\alpha$
Pitsch (1962)–Petch (1953)	$[100]_\theta$ 2.6° from $[31\bar{1}]_\alpha$ $[010]_\theta$ 2.6° from $[131]_\alpha$ $(001)_\theta \| (\bar{2}\bar{1}5)_\alpha$	$(001)_\theta \| (\bar{2}\bar{1}5)_\alpha$

either the Bagaryatsky or Pitsch–Petch orientation relationships have not yet been determined.

Zhou and Shiflet (1992) demonstrated that each orientation relationship possesses a unique interlamellar atomic habit plane (Table 1). Each of these habit planes is consistently maintained under all heat-treatment conditions, whether between lamellae in a single colony or between lamellae in different grains. Even lamellae that are curved preserve their habit planes listed in Table 1 for each orientation relationship. Curvature obtains with the addition of steps at the interlamellar boundaries, the terraces of which maintain the atomic matching.

4. *Edgewise Growth of Pearlite*

Pearlite has been shown to consist of two interwoven crystals of ferrite and cementite (Hillert 1962). A pearlite colony increases its volume by edgewise growth. Sideways growth occurs by lamella branching; no new crystals are created. Growth occurs by diffusion of the carbon to the advancing plate edges. The pearlite growth interphase boundary is historically considered the most likely to be crystallographically disordered or incoherent. There have been several studies of pearlite where the orientation relationship of a pearlite colony with the austenite grain it is growing into has been characterized, with the finding that no systematic orientation relationship occurred. However, Hackney and Shiflet (1987a) have shown that the absence of a reproducible orientation relationship does not preclude a partially coherent interphase boundary. They emphasized the role of crystallography in the growth of pearlite and developed an integrated mechanism of growth whereby the ledges at the pearlite/austenite growth interface also provide a means by which the position of the ferrite/cementite lamellar interface can be altered.

Pearlite growth kinetics at a given temperature is generally considered to be the same for all nodules and independent of time when occurring in a two-phase region. More careful measurements have shown that the growth rate at a given temperature is a function of time and can be different from nodule to nodule and from colony to colony in the same nodule. This is probably owing to the role of the nonreproducible, but still partially coherent interface with the austenite grain that is being consumed.

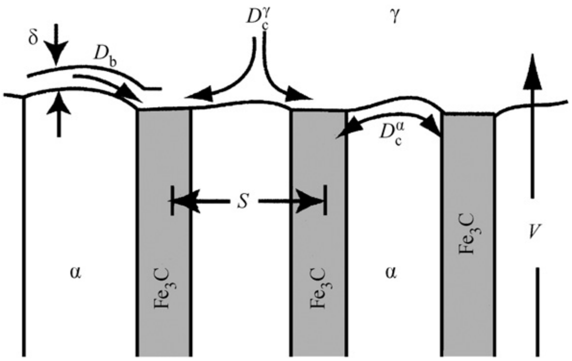


Figure 3
Schematic diagram of a pearlite colony with various growth parameters defined.

An edgewise growing colony of pearlite with interlamellar spacing S is shown in Fig. 3. For the transformation to progress, carbon redistribution from ferrite to cementite must occur for carbon-rich cementite (6.67% carbon) and carbon-depleted ferrite (<0.03% carbon) to grow. As shown in Fig. 3 a number of diffusion paths are available for carbon. Solute can redistribute in the austenite, in the ferrite, and along the interphase boundaries between pearlite and austenite. Volume diffusion in austenite appears to govern pearlite growth; however, contributions from diffusion through other diffusion paths cannot yet be ruled out and are likely to contribute. Carbon diffusion through ferrite may occur, as its diffusivity is much higher than in austenite. Carbon diffusion along the reaction interface can become substantial since diffusivity along an interface is usually higher than in the bulk, particularly at low temperatures. Figure 3 also highlights the important parameters associated with modeling pearlite growth, including the edgewise velocity, V, interlamellar spacing, S, and interphase boundary thickness, δ. Growth data for near-eutectoid Fe–C assuming control by diffusion of carbon in austenite have been measured but significant deviations between the experimental data and calculated growth rates exist at all temperatures, especially at lower ones. This suggests that carbon diffusion along pearlite/austenite boundaries significantly contributes to growth kinetics.

The widely used Zener–Hillert (Zener 1946, Hillert 1957) equation for pearlite growth is based on the assumptions that the supersaturated austenite is in local equilibrium with the constituent phases of pearlite at the reaction front, and the volume diffusion of carbon in austenite is rate controlling. This equation is derived from the flux balance of carbon required for the reaction and from an accounting made for the influence of lamellar interfacial energy on the driving force for growth. The interlamellar spacing of pearlite is thought to be constant at fixed temperatures and decreases markedly with decreasing temperature. An additional consideration is that if the lamellar spacing is sufficiently fine, S_c, then all the available energy is tied up in the interfaces, halting growth, thus yielding:

$$G = \frac{2D_c^\gamma}{f^\alpha f^\theta} \cdot \frac{x_\gamma^{\gamma\alpha} - x_\gamma^{\gamma\theta}}{x_\theta^{\theta\alpha} - x_\alpha^{\alpha\theta}} \cdot \frac{1}{S} \cdot \left(1 - \frac{S_c}{S}\right) \tag{1}$$

where D_c^γ is the volume diffusivity of carbon in austenite, the various x_i are defined in Fig. 2, and the fractions of the two phases, f^α, f^θ, are based on the lever rule. Equation (1) does not provide a unique solution because there are two unknown parameters, G and S, at a given undercooling. Attempts to specify S have been made by means of *ad hoc* optimization conditions, most notably by Zener (1946). Optimization theories do predict that undercooling and lamellar spacing are related by $S \cdot \Delta T = \text{constant}$, which for a eutectoid alloy have been shown to apply at small ΔT (about 60 °C) below the eutectoid temperature. Applying Zener's maximum growth theory gives the optimum pearlite spacing to be twice the critical spacing, $S = 2S_c$, and inserting into Eqn. (1) gives a growth rate proportional to S^{-1}. Boundary diffusion-controlled pearlite growth has also been considered (Hillert 1969) and follows a similar relationship where growth is now proportional to S^{-2} and is maximized when $S = 3S_c/2$. More rigorous derivations have been given by Hillert (1972).

5. *Ternary Fe–C–X Pearlite*

The addition of a second solute element to Fe–C invariably affects the formation of pearlite. Substitutional element additions alter the activity of carbon. Also, the partitioning of X between ferrite and cementite has a dramatic effect on growth. Indeed, improvement of hardenability is obtained by addition of elements that reduce pearlite formation kinetics, e.g., molybdenum. At small undercoolings (high temperatures) partitioning of the substitutional element can occur. However, as the reaction temperature is lowered diffusion of X becomes increasingly sluggish. There are two different types of partitioning behavior. The first is when the alloying element partitions preferentially to either pearlitic cementite (chromium, molybdenum) or pearlitic ferrite (silicon) at the transformation front over the full temperature range of pearlite formation, although the extent of partitioning decreases progressively as the temperature falls. The second is the alloying addition, e.g. manganese, partitions preferentially at the transformation front at high temperatures but the extent of partitioning decreases until a temperature can be identified below which there is no partitioning of the alloying element at the growth front. Ridley has been involved in many of these experimental studies and has given an extensive review (Ridley 1984). Another effect of addition of elements such as manganese, depending on composition and reaction temperature, is that pearlite can grow in either a two-phase region where all the austenite should be consumed by pearlite, or a three-phase region where austenite coexists with pearlite at equilibrium. The consequences of growth in a three-phase field (ferrite, cementite, austenite) are increasing lamellar spacing and a continuously changing growth rate with time during isothermal treatments. Cahn and Hagel (1962) call this divergent pearlite and Hillert (1982) has given a thermodynamic description of this observation. Another possible occurrence in ternary alloys is the growth of pearlite in a two-phase austenite + cementite phase field above the eutectoid temperature in hypereutectoid alloys. This results from the slow partitioning of manganese and local equilibrium conditions applied to carbon, which allow ferrite to form. The addition of elements that form carbides, such as vanadium and chromium, can lead to parallel sheets of carbides within the ferrite lamella. These interphase boundary carbides map out the position of the pearlite growth interface. This microstructure has been interpreted by Zhou and Shiflet (1991) to be the result of growth control by means of ledges.

6. *Summary*

Much has been learned about pearlite in steels since the early twentieth century. Further understanding of this very unique phase transformation will greatly benefit with the improvement of analytical and high-resolution microscopes and the greater access to thermodynamic databases. Much more needs to be learned to ascertain correctly the roles of crystallography, solute partitioning paths, and concomitant precipitation within pearlite in the progress of the pearlite reaction.

Bibliography

Bagaryatsky Y A 1950 Veroyatnue mechanezm raspada martenseeta. *Dokl. Akad. Nauk. SSSR* **73**, 1161–4

Cahn J W, Hagel W C 1962 Theory of the pearlite reaction. In: Zackay V F, Aaronson H I (eds.) *Decomposition of Austenite by Diffusional Processes*. Interscience, New York, pp. 131–92

Hackney S A, Shiflet G J 1987a The pearlite:austenite growth interface. *Acta Metall.* **35**, 1007–17

Hackney S A, Shiflet G J 1987b Pearlite growth mechanism. *Acta Metall.* **35**, 1019–28

Hillert M 1957 The role of interfacial energy during solid state phase transformations. *Jerkont. Ann.* **141**, 757–89

Hillert M 1962 The formation of pearlite. In: Zackay V F, Aaronson H I (eds.) *Decomposition of Austenite by Diffusional Processes*. Interscience, New York, pp. 197–237

Hillert M 1969 The role of interfaces in phase transformations. In: Nicholson R B, Christian J W, Greenwood G W, Nutting J, Smallman R E (eds.) *The Mechanism of Phase Transformations in Crystalline Solids,* Monograph 33. Institute of Metals, London, pp. 231–47

Hillert M 1972 On theories of growth during discontinuous precipitation. *Metall. Trans.* **3**, 2729–41

Hillert M 1982 An analysis of the effect of alloying elements on the pearlite reaction. In: Aaronson H I, Laughlin D E, Sekerka R F, Wayman C M (eds.) *Proc. Int. Conf. on Solid–Solid Phase Transformations*. TMS-AIME, Warrendale, OH, pp. 789–806

Isaichev I V 1947 Orientation between cementite and ferrite. *Zh. Tekh. Fiz.* **17**, 835–8

Mangan M A, Shiflet G J 1999 The Pitsch–Petch orientation relationship in ferrous pearlite at small undercooling. *Metall. Trans.* **30A**, 2767–81

Mehl R F, Hagel W C 1956 The austenite:pearlite reaction. *Prog. Metal Phys.* **6**, 74–134

Petch N J 1953 The orientation relationships between cementite and α-iron. *Acta Cryst.* **6**, 96

Pitsch W 1962 Der Orientierungszusammenhang zwischen Zementit und Ferrit im Perlit. *Acta Metall.* **10**, 79–80

Ridley N 1984 The pearlite transformation. In: Marder A R, Goldstein J I (eds.) *Proc. Int. Conf. Phase Transformations in Ferrous Alloys*. TMS-AIME, Warrendale, OH, pp. 201–36

Sorby H C 1886 On the application of very high powers to the study of the microscopic structure of steel. *JISI* **1**, 140–7

Zener C 1946 Kinetics of the decomposition of austenite. *Trans. AIME* **167**, 550–83

Zhou D S, Shiflet G J 1991 Interfacial steps and growth mechanism in ferrous pearlite. *Metall. Trans.* **22A**, 1349–65

Zhou D S, Shiflet G J 1992 Ferrite: cementite crystallography in pearlite. *Metall. Trans.* **23A**, 1259–69

G. J. Shiflet
University of Virginia, Charlottesville, Virginia, USA

Phase Diagrams

Properties of materials are related to their microstructure. In turn, the microstructure is related to the constitution, which states the number of phases present, their proportion, and their composition. To determine the constitution we need to know the phase diagram.

Typically, in a solid, microstructure consists of grains of phases that abut against one another. These grains may consist of only one phase (a single-phase microstructure), or they may be grains of many phases (a polyphase microstructure).

An assessment of the overall properties of a material may proceed by determining first the constitution of the phases involved, then combining this with the knowledge of their specific properties (i.e., crystal structures, physical properties, etc.) and, finally and very importantly, incorporating the knowledge of the *distribution* of the phases in the microstructure, (i.e., the shapes and sizes of the grains, how they fit together, and how they interact, etc.). In this process, the knowledge of the constitution is an important initial step.

The information that allows us to predict the constitution from a graphical plot of phases that may be involved is known as a phase diagram. Hence, constitution is a function of what kind of a phase diagram is used for a given material. The independent state variables that control the phases present in a given phase diagram are the number of components (i.e., elements) involved, the chemical composition (i.e., the proportion of the components in the material), the temperature, the external pressure, magnetic field, electrostatic field, etc. Simple phase diagrams generally involve only two state variables, for example the temperature and composition. If one is plotted vertically and the other horizontally, a graphical representation is obtained, showing the distribution of the various possible phase fields allowed by the two variables, as will be shown below.

1. Phases

All materials exist in gaseous, liquid, or solid form, usually referred to as a phase, depending on the conditions of state (i.e., the magnitude of state variables involved). The term "phase" refers to that region of space occupied by a physically homogeneous material.

In a phase diagram, each single-phase field is depicted graphically and is usually given a single label. Engineers often find it convenient to use this label to refer to all the material having a constitution related to this field, regardless of how much the physical properties of the materials may be continuously changing from one part of the field to another. This means that in engineering practice the distinction between the terms "phase" and "phase field" is often neglected, and all materials having the same phase name are referred to as the same phase.

2. Equilibrium

There are three states of equilibrium: stable, metastable, and unstable. These three conditions are illustrated schematically in a structural sense in Fig. 1.

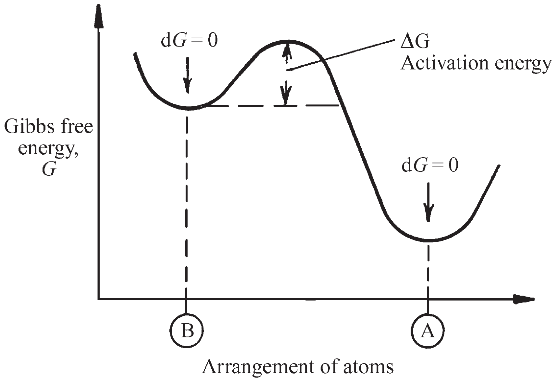


Figure 1
Free energy as a function of the arrangements of atoms in phases. A: stable configuration, B: metastable configuration (after Porter and Easterling 1992).

Stable equilibrium exists when the object is in its lowest energy condition; metastable equilibrium exists when additional energy (ΔG) must be introduced before the object can reach true stability; unstable equilibrium exists when no additional energy is needed before reaching metastability or stability. Although true stable equilibrium conditions seldom exist in real everyday materials, the study of equilibrium systems is extremely valuable, because it constitutes a limiting condition from which actual conditions can be estimated.

3. Thermodynamic Considerations

The state variables that are selected to control a given system determine the free energy values of all the possible phases that can exist in that system for all possible combinations of the variables (all the state points in a state diagram plot). The phases that are actually recorded in a phase diagram represent the lowest value of the free energy at each state point.

4. The Gibbs Free Energy

The Gibbs free energy of a system may be defined as:

$$G = H - TS \tag{1}$$

where H is the enthalpy, T is the absolute temperature, and S is the entropy. Enthalpy is the "heat content," and is therefore related to the specific heat of the given phase.

$$H = E + PV \tag{2}$$

where E is the total "bonding" energy, P is the pressure, and V is the volume. A system is in equilibrium when it shows no tendency to change, or

$$\mathrm{d}G = 0 \tag{3}$$

The immediate states for which $\mathrm{d}G \neq 0$ are unstable and are only realized for limited periods in practice, unless the temperature is very low and atoms cannot move. If, as the result of thermal fluctuations, the atoms become arranged in an intermediate state, they will rearrange again into a free energy minimum. If by a change of temperature or pressure, for example, a system is moved from a stable to a metastable state, it will, given time, transform to a new equilibrium state later.

Graphite and diamond at room temperature and pressure are examples of stable and metastable equilibrium states. Given time therefore, diamond under these conditions will transform into graphite. Similarly, the cementite constituent in steels (Fe_3C) is metastable and will transform to graphite and iron.

Any transformation that results in a decrease in Gibbs free energy is possible if the kinetics are favorable. Therefore a necessary criterion for any phase transformation is

$$\Delta G = G_2 - G_1 < 0 \tag{4}$$

where G_1 and G_2 are the free energies of the initial and final states respectively. The transformation need not go directly to the stable equilibrium state, but can pass through a whole series of intermediate metastable states. With extremely rapid freezing of liquids, even thermodynamically unstable structures (such as amorphous metallic "glasses") can be produced.

5. Systems

A physical system consists of a set of components (i.e., substances, or a single component) that is isolated from its surroundings, a concept used to facilitate study of the effects of conditions of state. "Isolated" means that there is no interchange of mass between the substance and its surroundings. The components in alloy systems, for example, may be two metals, such as copper and zinc; a metal and a nonmetal, such as iron and carbon; a metal and an intermetallic compound, such as iron and cementite; or several metals, such as aluminum, titanium, and vanadium. Components are almost always elements, but sometimes they can be designated as stoichiometric chemical compounds. These components comprise the system and should not be confused with the various phases found within the system.

5.1 System Components

Phase diagrams and the systems they describe are often classified and named for the number (in Latin) of components in the system illustrated in Table 1.

Table 1
Classification of phase diagrams and systems.

Number of components	Name of system or diagram
1	Unary
2	Binary
3	Ternary
4	Quaternary
5	Quinary, etc.

Source: Massalski (1998)

Table 2
Possible number of systems (n = number of elements (say 90); m = number of elements in a system).

	$\frac{n!}{m!(n-m)!}$
Binary	$\frac{90!}{2!88!} = 4005$
Ternary	$\frac{90!}{3!87!} = 117480$
Quaternary	$\frac{90!}{4!86!} = 2555190$

Clearly, as the number of components increases, so does the number of possible systems (Table 2).

This table shows that even if only four components are chosen from a group of, say, 90 elements, the possible number of combinations runs into millions, which is a warning that the phase diagrams of quaternary systems are a long way from being determined, unless they can be calculated via a computer program using estimates of free energies.

6. Phase Diagrams

These are graphical plots that were devised to show the relationships between the various phases that appear within the system under equilibrium conditions. As such, the diagrams are variously called constitutional diagrams, equilibrium diagrams, or phase diagrams. A single-component phase diagram can be simply a one- or two-dimensional plot, showing the phase changes in the given component (for example in iron) as temperature and/or pressure changes. Most diagrams, however, are two- or three-dimensional plots describing the phase relationships in systems made up of two of more components, and these usually contain fields (areas) consisting of mixed-phase fields, as well as single-phase fields. Some typical examples are shown below.

7. Phase Rule

The phase rule, first announced by J. Willard Gibbs in 1876, relates the physical state of a mixture to the number of constituents in the system and to its conditions. It was also Gibbs who first called each homogeneous region in a system by the term "phase." When pressure and temperature are the state variables, the rule can be written as follows:

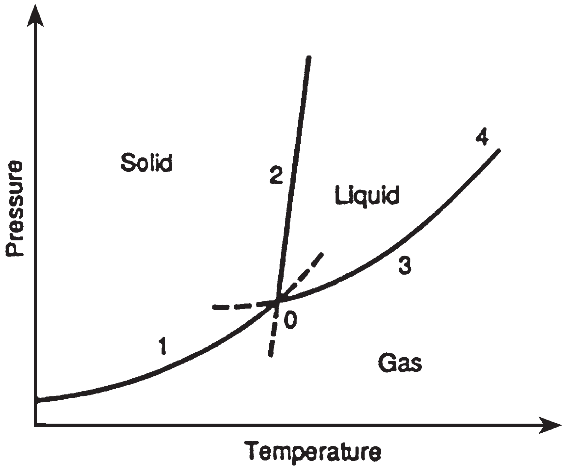


Figure 2
Schematic pressure-temperature phase diagram (after Baker 1992).

$$f = c - p + 2 \tag{5}$$

where f is the number of independent variables (called degrees of freedom), c is the number of components, and p is the number of state phases in the system. The Gibbs phase rule applies to all states of matter (solid, liquid, and gaseous), but when the pressure is constant, the rule reduces to:

$$f = c - p + 1 \tag{6}$$

8. Unary Systems and Phase Diagrams

8.1 Invariant Equilibrium

According to the phase rule, three phases can exist in a stable equilibrium only at a single point of a unary diagram ($f = 1 - 3 + 2 = 0$). This limitation is illustrated as point 0 in the hypothetical unary pressure–temperature (PT) diagram shown in Fig. 2. In this diagram, the three states (or phases)—solid, liquid, and gas—are represented by the three correspondingly labeled fields. Stable equilibrium between any two phases occurs along their mutual boundary, and invariant equilibrium among all three phases occurs at the so-called triple point, 0, where the three boundaries intersect. This point also is called an invariant point because, at that location on the diagram, all externally controllable factors are fixed (no degrees of freedom). At this point, all three states

Table 3
Stable equilibria for binary systems.

Number of components	Number of phases	Degrees of freedom	Equilibrium
2	3	0	Invariant
2	2	1	Univariant
2	1	1	Bivariant

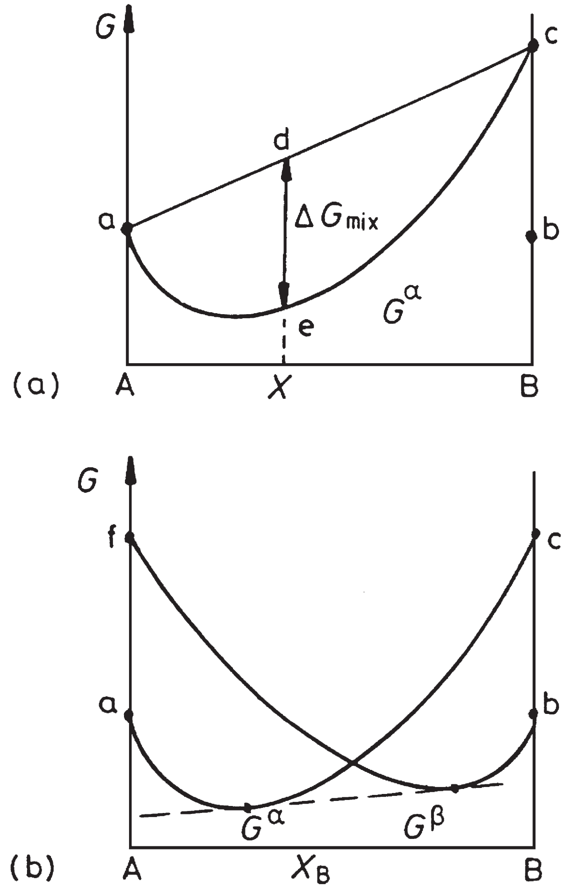


Figure 3
(a) Molar free energy curve for the α phase. (b) Molar free energy curves for α and β phases (after Porter and Easterling 1992).

(phases) are in equilibrium, but any changes in pressure and/or temperature will cause one or two of the states (phases) to disappear.

8.2 Univariant Equilibrium

The phase rule says that stable equilibrium between two phases in a unary system allows one degree of freedom ($f = 1 - 2 + 2$). This condition, called univariant equilibrium, or monovariant equilibrium, is illustrated as lines 1, 2, and 3 separating the single-phase fields in Fig. 2. Either pressure or temperature may be freely selected, but not both. Once a pressure is selected, there is only one temperature that will satisfy equilibrium conditions and, conversely, the three curves that issue from a triple point are called triple curves: line 1, representing the reaction between the solid and the gas phases, is the sublimation curve; line 2 is the melting curve; and line 3 is the vaporization curve. The vaporization curve ends at point 4, called a critical point, where the physical distinction between the liquid and gas phases disappears.

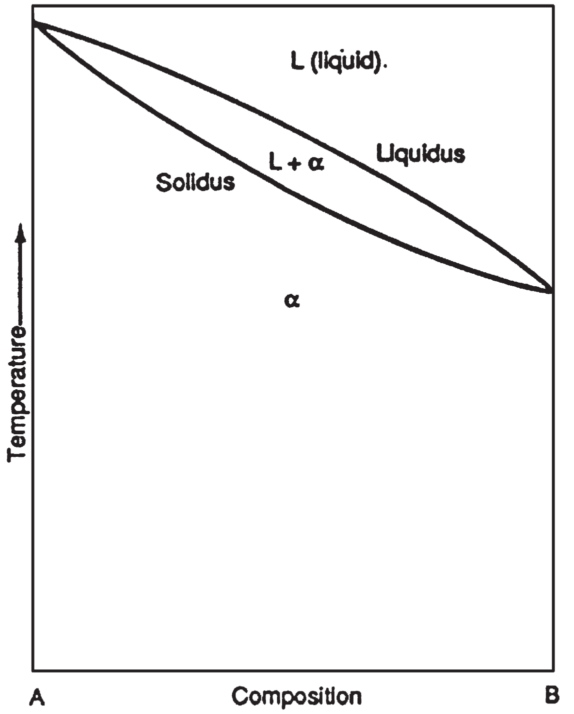


Figure 4
Schematic binary phase diagram showing miscibility in both the liquid and the solid states.

9. Bivariant Equilibrium

If both the pressure and temperature in a unary system are freely and arbitrarily selected, the situation corresponds to having two degrees of freedom, and the phase rule says that only one phase can exist in stable equilibrium ($p = 1 - 2 + 2$). This situation is called bivariant equilibrium. The stable equilibria for binary systems are summarized in Table 3.

10. Phase Stability

The extent of solid solubility of phases, the stability of phases, the temperature dependence of stability, and the choice of structures that are actually observed in phase diagrams are the result of competition among numerous possible structures that could be stable in a given system. This competition is based on the respective values of the Gibbs free energy of each competing phase and the variation of this energy with temperature, pressure, composition, and possibly other extensive parameters. In its simplest form, the free energy can be written as in Eqn. (1).

As is well known, numerous factors contribute to the H and S parameters. The major contribution to the entropy is from statistical mixing of atoms (ΔS_{mix}), but there can be additional contributions from vibrational effects (ΔS_{vib}), distribution of magnetic moments, clustering of atoms, and various long-range configurational effects. The main contributions to the enthalpy result from atomic mixing (ΔH_{mix}), which are in turn related to the interaction energies between neighboring and further distant atoms in a given structure and are based upon electronic, elastic, magnetic and vibrational effects. Much progress has been made in measuring, calculating, and predicting many such effects, and hence progress continues to be made in the evaluation of the related thermodynamic quantities and ultimately the phase diagrams.

In a binary system, the free energy G of a given phase typically has a downward parabolic form between the values for the two components (say A and B) as composition changes. This is illustrated schematically in Fig. 3(a). If the structures of the two components are different, two free energy curves may have to be considered at every temperature to determine lowest energies, as in Fig. 3(b). Such curves will become displaced as temperature changes, and at each temperature the lowest free energy configuration across the diagram will determine the form of the phase diagram at each composition. The well-known "tangent construction" is used for that purpose to determine the free energies when mixtures of phases result in lowest G.

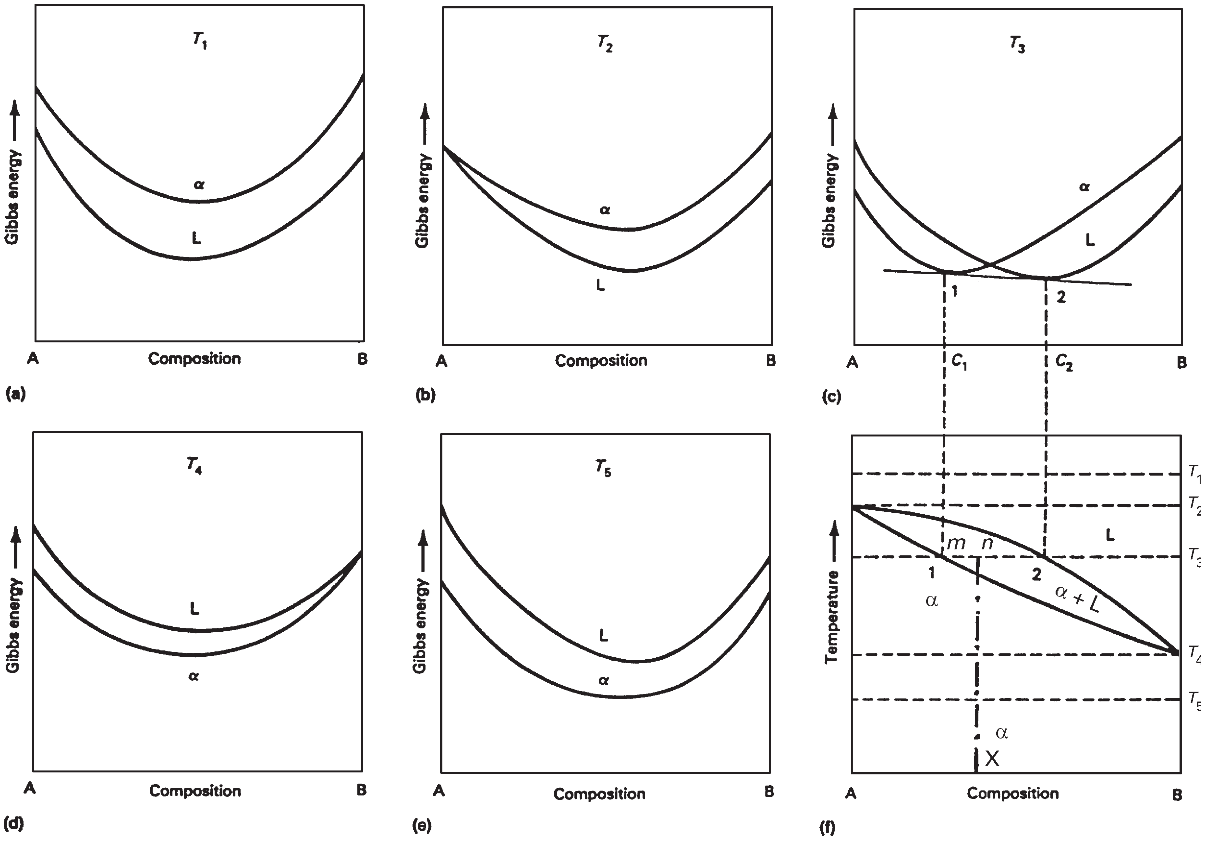


Figure 5
Derivation of a phase diagram with complete liquid and solid miscibility from free energy curves (after Prince 1966).

11. Features of Phase Diagrams

The areas (fields) in phase diagrams, and the positions and shapes of the points, lines, surfaces, and intersections in it, are controlled by thermodynamic principles and the thermodynamic properties of all of the phases that constitute the system.

11.1 Phase-field Rule

The phase-field rule specifies that at constant temperature and pressure, the number of phases in adjacent fields in a multicomponent diagram must differ by one.

12. Binary Diagrams

The study of systems of more than one component involves a study of solutions. The simplest type of a multicomponent system is a binary, and the least complex structure in such a system is a single solution. Some of the aspects of single-phase binary systems will now be considered.

12.1 Solid Solutions

We shall consider here only two state variables: temperature and composition. The replacement of nickel atoms by copper atoms on the lattice of nickel is an example of a substitutional solid solution. Nickel can dissolve copper at all proportions, providing an example of complete solid solubility. This is possible only if the sizes of the atoms differ by no more than about 15%. There are many examples of restricted mutual solid solubilities, even between elements with similar crystal structures and atom sizes. When a continuous solid solution occurs in a binary system, the phase diagram usually has the general appearance of that shown in Fig. 4. The diagram consists of two single-phase fields separated by a two-phase field. The boundary between the liquid field and the two-phase field in Fig. 4 is called the liquidus; that between the two-phase field and the solid field is called the solidus. In general, the liquidus is the locus of points in a phase diagram representing the temperatures at which alloys of the various compositions of the system begin to freeze on cooling, or finish melting on heating; a solidus is the locus of points

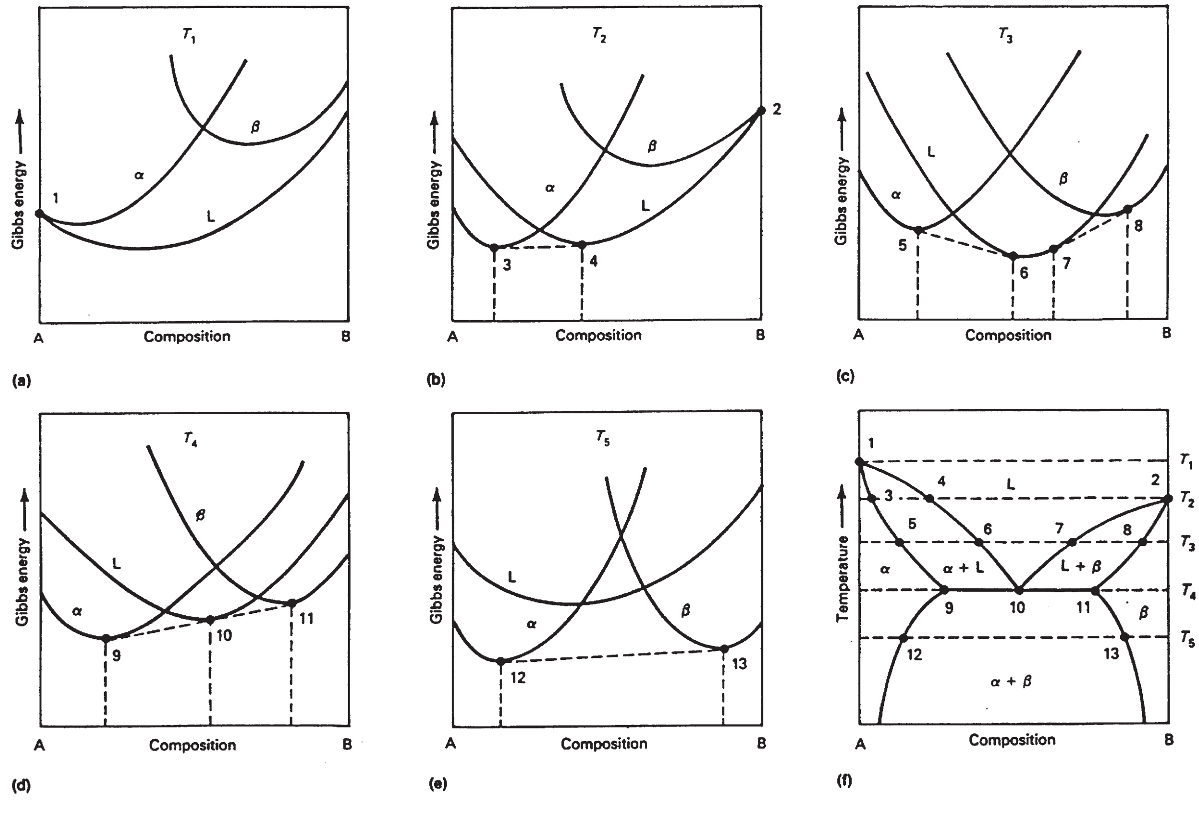

(a)
T1
β
α
1
L
Gibbs energy
A
Composition
B
(b)
T2
2
β
α
L
3
4
Gibbs energy
A
Composition
B
(c)
T3
L
β
α
5
6
7
8
Gibbs energy
A
Composition
B
(d)
T4
β
L
α
9
10
11
Gibbs energy
A
Composition
B
(e)
T5
L
α
β
12
13
Gibbs energy
A
Composition
B
(f)
1
2
3
4
5
6
7
8
9
10
11
12
13
L
α
α + L
L + β
β
α + β
T1
T2
T3
T4
T5
Temperature
A
Composition
B

Figure 6
Derivation of a eutectic phase diagram involving free energies of α, β and liquid phases (after Gordon 1968).

representing the temperatures at which the various alloys finish freezing on cooling, or begin melting on heating. The phases in equilibrium across the two phase-field in Fig. 4. (the liquid and the solid solutions) are called conjugate phases.

12.2 Free Energies and Binary Phase Diagrams

We can now illustrate a few examples of binary phase diagrams. The simplest case is a system which is completely, if not ideally, miscible in both the liquid and solid states. The free energies $G_{(L)}$ and $G_{(S)}$ are both parabolas, the relative positions of which change with temperature as shown in Fig. 5. If the melting points of the pure metals are different, the S ($= \alpha$) becomes tilted. Intersections and tangent constructions then become possible, yielding the well-known phase diagram with a liquid and solid showing complete miscibility, as in Fig. 4. The Ti–Zr phase diagram (Fig. 7(a)) is a good example of this behavior.

12.3 Constitution and Lever Rule

Figure 5 also provides the basis for understanding how the constitution (mentioned in Sect. 1) can be derived for any given alloy. Figure 5(c) illustrates that at temperature T_3 the lowest free energy path between A and B will be along the G curve for the α-phase (from A to C_2), then along the tangent between C_1 and C_2, and finally, along the G curve for the liquid between C_2 and B. The tangent contacts 1 and 2 represent the situation when the alloy is

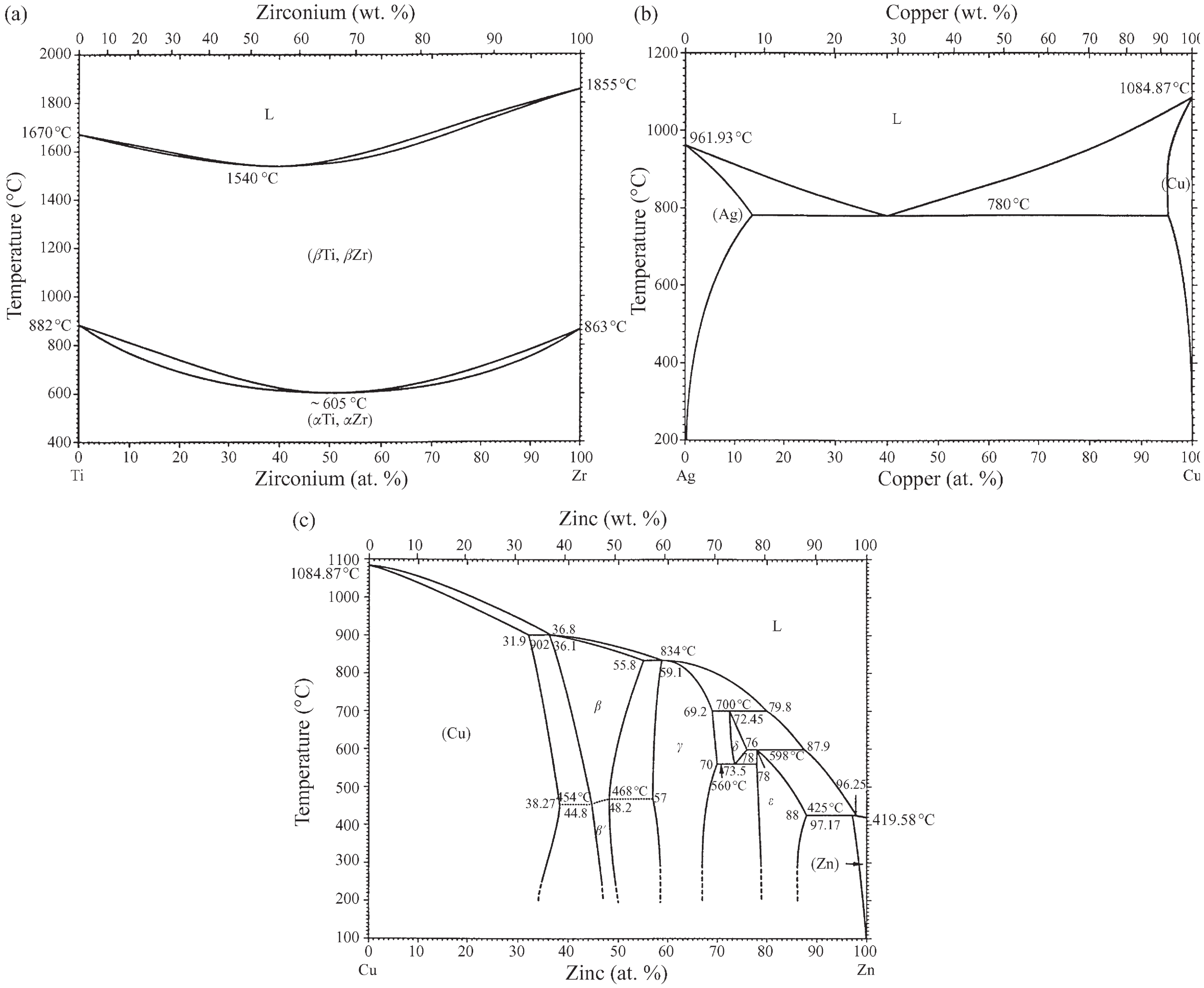


Figure 7
Assessed phase diagrams: (a) Ti–Zr, (b) Cu–Ag, (c) Cu–Zn (after Massalski 1990, ASM/NIST Phase Evaluation Program).

two-phase (α of composition C_1 and L of composition C_2), and this is the basis of the phase diagram shown in Fig. 5(f), at temperature T_3. The knowledge of constitution stipulates that we also determine proportions of the two phases present in equilibrium for any alloy within the two-phase field, and this is done by employing the lever rule. Thus, for an alloy of composition X at T_3, the proportion is given by the ratio of the two segments m and n as follows:

$$\alpha/\mathrm{L} = n/m \tag{7}$$

Other forms of phase diagram may be derived in the same way. For example, a eutectic form of the phase diagram is obtained if two free energy curves are involved for the solid phases, as shown in Fig. 3(b); or if the solid phase free energy curve is continuous, but has a positive "hump" in the middle of the composition range, typically because of atomic size difference effects.

This is illustrated schematically in Fig. 6. The Cu–Ag phase diagram (Fig. 7(b)) is a good actual example of a binary eutectic. Constitution of any alloy in this system, for any state point, could be derived by the same procedure as mentioned above for alloy X in Fig. 5(f).

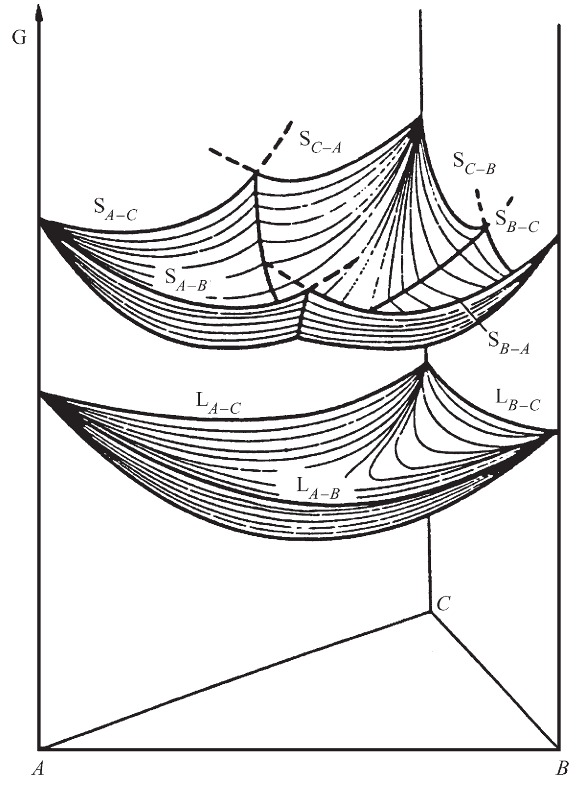


Figure 8
Free energies of L and three solid phases S of a ternary system (after Haasen 1996).

12.4 More Complex Diagrams

More complex phase diagrams may involve the presence of one or more "intermediate phases," which have the lowest free energies in the system at some temperatures and compositions. A good example is the Cu–Zn binary phase diagram, shown in Fig. 7(c), which involves two terminal phases, α and η, and five intermediate phases: α, β, γ, δ, and ε.

With a little practice, a set of hypothetical free energy curves could be constructed for all the phases, as well as their respective displacements with temperature, such that the overall free energy trends would yield the observed diagram. This serves to emphasize the fact that many selected sets of free energy curves can be postulated that would reproduce an actual phase diagram. Only experimental measurements of some of the thermodynamic quantities representing the actual phases can confirm which free energy curves are really valid and represent the actual free energy situations.

13. Ternary (and Higher) Component Phase Diagrams

When a third component is added to a binary system, the graphic representation of the equilibrium conditions in two dimensions becomes too complicated. One option is to add a third composition dimension, by using a triangular base with a vertical dimension

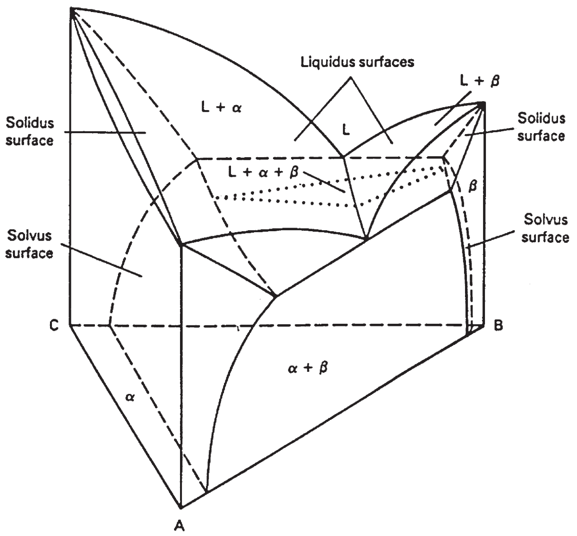


Figure 9
Ternary phase diagram showing three-phase equilibrium (after Rhines 1956).

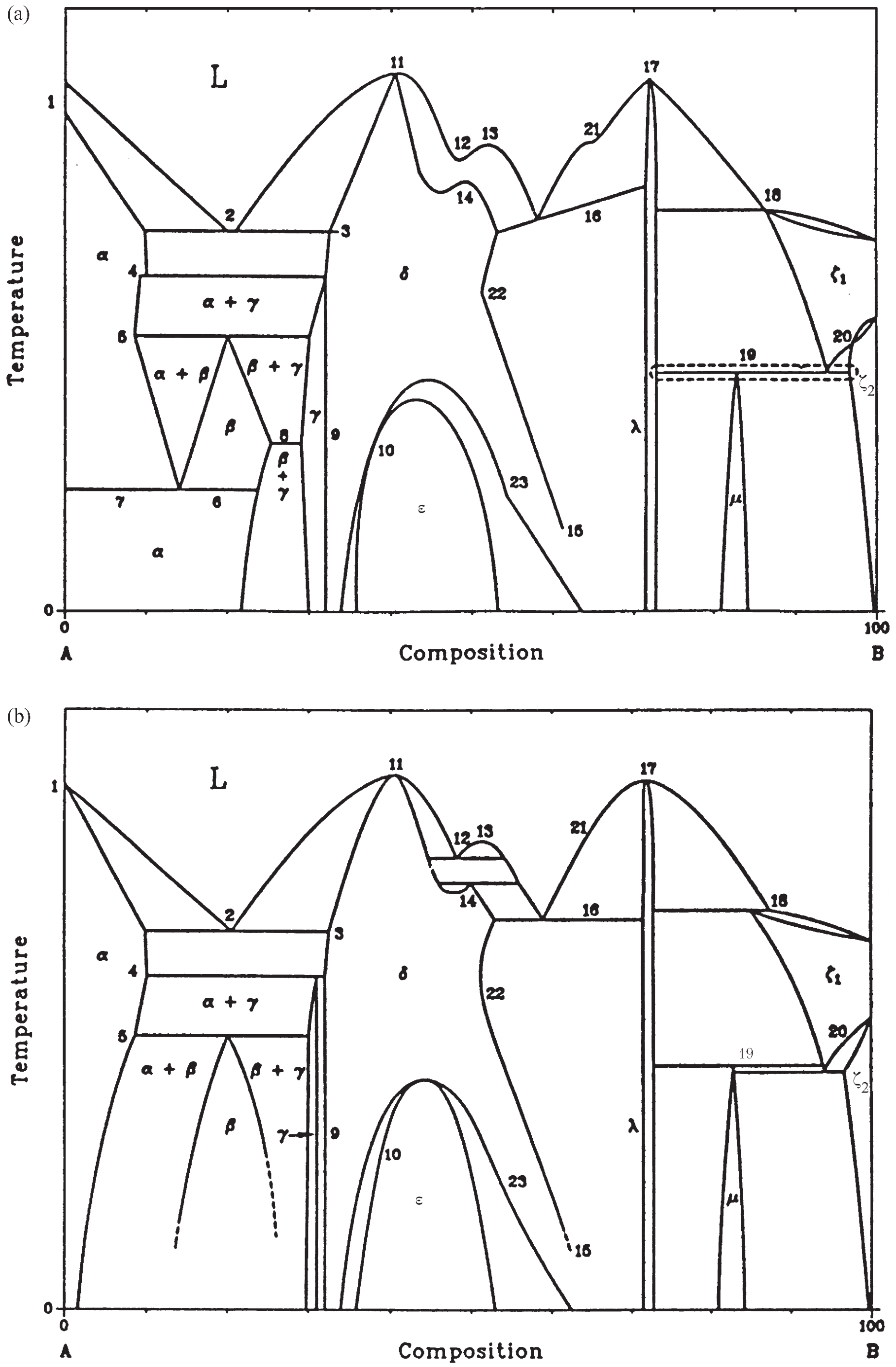

(a)
L
1
11
17
12
13
21
14
2
3
16
18
α
4
δ
ζ1
α + γ
22
5
20
19
α + β
β + γ
ζ2
γ
β
8
9
10
λ
β
+
γ
23
7
6
ε
μ
15
α
Temperature
0
0
100
A
Composition
B
(b)
L
1
11
17
12
13
21
14
2
16
18
3
α
4
δ
ζ1
α + γ
22
5
20
19
α + β
β + γ
ζ2
β
γ
9
λ
10
23
ε
μ
15
Temperature
0
0
100
A
Composition
B

Figure 10
(a) Hypothetical binary phase diagram showing many typical errors of construction. (b) Error-free version of the phase diagram shown in (a) (after Okamoto and Massalski 1991).

for temperature. Free energy curves now become free energy humps, lines become surfaces, etc. This is illustrated in Fig. 8. A corresponding phase diagram may be as that illustrated schematically in Fig. 9. Even more complex graphics have been proposed for higher systems, but generally only selected portions of the actual diagrams can be represented. Further details will be found in the cited literature.

14. *Phase Diagram Errors*

Because phase diagrams are an expression of thermodynamic conditions, it is necessary to ensure that the constructed graphics are thermodynamically correct and compatible. Unfortunately, this is frequently not the case, and hence many efforts have been made in the past, and continue to be made, to evaluate critically the existing literature on phase diagrams (see *Phase Diagrams: Data Compilation*).

An example of a hypothetical complex phase diagram involving 23 errors is shown in Fig. 10(a), and an error-free version of the same diagram is shown in Fig. 10(b). Some errors are quite obvious, even to beginners, but the reader will need to consult the cited literature (Okamoto and Massalski 1991) for a detailed discussion of all the depicted errors and their correction.

15. *Determination of Phase Diagrams*

A very large literature exists on the various methods for determining phase diagrams. Historically, many of the well-known diagrams were determined experimentally, long before the obvious connection between phase thermodynamics and phase diagrams was fully realized. Experimental techniques involve metallography, cooling and heating thermal arrest studies (DSC techniques), x-ray measurements, magnetic measurements, dilatometry, electron microprobe, etc., to mention only a few.

Parallel with this activity, an extensive literature has grown on the measurements of the thermodynamic properties of elements, compounds, and phases, which can then be used to derive the respective free energies and their changes with different state variables. Once free energies are known, phase diagrams can be derived via computer calculation and modeling.

It is also possible to perform no experimental work at all (either metallurgical or thermodynamic), and to use computer programs that will simulate and optimize free energies from assumed models of different structures, and their bonding energies, and then to derive phase diagrams entirely by calculation. It is obvious that many opportunities and pitfalls exist in all of these approaches.

See also: Phase Diagrams: Data Compilation

Bibliography

Baker H 1992 Alloy phase diagrams. In: *ASM Handbook*, Vol. 3. ASM International, Materials Park, OH, pp. 1–30

Gordon P 1968 *Principles of Phase Diagrams in Materials Science*. McGraw-Hill, New York

Haasen P 1996 *Physical Metallurgy*, 3rd edn. Cambridge University Press, Cambridge

Massalski T B 1984 Phase diagrams in materials science. *ASM News,* **20**, 8–9

Massalski T B 1998 *Binary Phase Diagrams*. ASM/NIST Phase Evaluation Program 98. ASM International, Materials Park, OH

Okamoto H, Massalski T B 1991 Thermodynamically improbable phase diagrams. *J. Phase Equilibria* **12**, 148–68

Porter D A, Easterling K E 1992 *Phase Transformations in Metals and Alloys*, 2nd edn. Chapman & Hall, London

Prince A 1966 *Alloy Phase Equilibria*. Elsevier, London

Rhines F N 1956 *Phase Diagrams in Metallurgy*. McGraw-Hill, New York

T. B. Massalski
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Phase Diagrams: Data Compilation

One of the characterizing features of a given alloy is the set of phases that coexist at a given temperature. When exposed to a different temperature, phases may dissolve, others may form, and the relative amounts of the phases present may change. The results can be seen in micrographs. This rearranging may involve intermediate phase configurations that are unstable over time, or metastable. The system will continue to form finally stable phase configurations, driven by thermodynamic potentials, for that alloy composition. Phase diagrams commonly reflect the thermodynamic stable equilibrium state.

Phase diagrams reveal the sets of phase configurations that coexist, thermodynamically stable or metastable, when temperature, pressure, or the concentrations of the constituent elements vary.

To know for a given alloy the phases that one can expect is a basic requirement for understanding its metallurgical processes and applications. For the development of new candidate materials it is vital to know at the earliest possible time the phases that might be formed, transform, or dissolve. To know how the phase configurations change and the conditions under which the phase reactions occur reveals possible processing routes to establish (or to avoid) specific phase configurations.

The controlling parameters are the temperature, pressure, and the selection of elements that interact. They control the alloying process and the service life of a component, where the working materials in the environment interact with the component as a function of pressure and temperature. Targeted development of new "high-tech" materials without an idea of the possible phase configurations is commercially and technically almost impossible, because of the number of independent variables.

1. The Number of Possibilities

The number of possible alloy systems, i.e., binary, ternary, multicomponent, is extremely large. If one assumes that 80 elements are sensible candidates for engineering materials, the numbers of possible element combinations yield something like 10^{23} possibilities, the variation of concentrations within each element combination not being considered. Modern technical alloys may contain up to 12 elements, and knowledge about possible phase configurations is scarce. Compared to these numbers, the number of at least partially examined systems is very small. For the important group of ternary systems only a little more than 15% of the candidate systems have ever been studied experimentally.

Nevertheless, the amounts of known data and diagrams are so large and their intercorrelation so complex that only coordinated international efforts can organize evaluation and access.

2. APDIC, the Alloy Phase Diagram International Commission

APDIC was set up in 1986 to safeguard the quality of phase diagram evaluations and to provide globally the best possible coordination of the major phase diagram projects. Under the auspices of APDIC, the Binary Evaluation Program was completed under the auspices of ASM International (see Sect. 7 (Massalski)). The MSIT Higher Order Evaluation Program continues under the auspices of MSI, Materials Science International Services, Stuttgart, Germany (see Sect. 7 (MSIT Workplace)). The members of APDIC are:

- Argentine Committee for Phase Diagrams, Argentina
- ASM International, USA
- Baikov Institute of Metallurgy, Russia
- Brazilian Committee for Phase Diagrams of Materials, Brazil
- Canadian Phase Diagram Research Association, Canada
- Chinese Phase Diagram Committee, China
- Department of Chemistry, University of Genoa, Italy
- Deutsche Gesellschaft für Materialkunde, Germany
- Groupe de Thermodynamique et Diagrammes de Phases, France
- Indian Institute of Metals, India
- Institute of Materials, UK
- Japanese Committee for Alloy Phase Diagrams, Japan
- Korean Committee of Thermodynamics and Phase Equilibria, Korea
- Materials Science International Services, Stuttgart, Germany
- Max-Planck-Institut für Metallforschung, Stuttgart, Germany
- Polish Academy of Sciences, Poland
- Royal Institute of Technology, Sweden
- Ukrainian Phase Diagrams and Thermodynamics Commission, Ukraine

The annual activity reports of APDIC members can be obtained from www.msiwp.com.

3. Thermodynamic Calculation of Phase Diagrams

In the past graphical representations were mainly used for discussing constitutional problems. Now, in addition, powerful simulation programs are available for calculation of phase diagrams (Calphad). These allow approximations even into concentration–temperature regions where no experimental data are available, provided the thermodynamic data are available. The leading group in generating such data is the Scientific Group Thermodata Europe (SGTE); application software packages such as Thermocalc, Chemsage, and Fact are used, and, as the latest, a user-friendly development called Pandat.

Owing to the large number of possible materials systems it is necessary that any materials development is supported by computer simulations to limit the number of experimental studies necessary.

Compilations of critically evaluated phase diagram data (see Sect.7 (Massalski, Okamoto, MSIT Workplace)) provide the information to validate and check the computer results against reality and can partially substitute missing thermodynamic data.

4. Phase Diagram Publications

Phase diagram data can be found in journals, monographs, compilations, and also in an electronic information system (Sect. 7 (MSIT Workplace)) that arranges all data on all alloy systems in terms of element combination. The information can be found in topical monographs such as 'Binary Magnesium Systems,' etc., journals such as *Journal of Alloys and Compounds, Zeitschrift für Metallkunde, Journal of Phase Equilibria*, etc., and compilations such as the Red Book series, 'Binary Phase Diagrams,' the Ternary Alloys series, etc.

It is worth pointing out that one third of the total of phase diagram publications, starting in 1900,

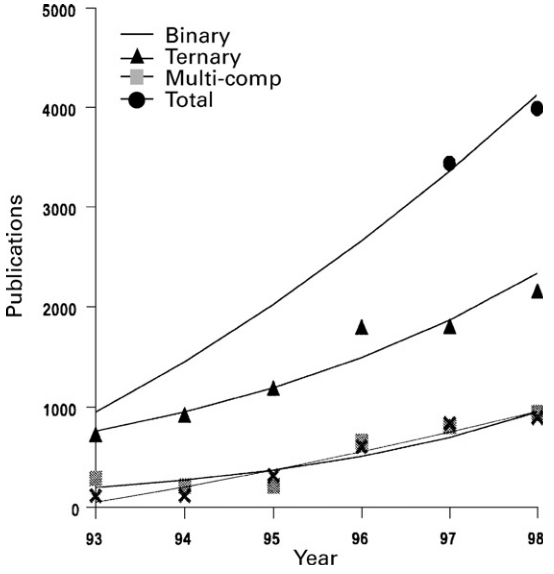


Figure 1
Phase diagrams and constitutional data in the literature.

originates from the former Soviet Union. To obtain the original experimental results one would have to monitor over 300 journals in order to stay informed. This is done by the Materials Science International Team; Fig. 1 shows the growing rate of data published, as retrieved from the MSIT Workplace.

5. *Critical Evaluation of Phase Diagrams and Related Data*

There is a wide range of experiments, distinctly different in kind, that can give clues to the constitution of alloys. These may be micrographic, thermal, or phenomenological in nature and as different as measurements of hardness, electrical resistivity, mass spectrometry, calorimetry, and microscopy.

Many experiments are done with a specific purpose in mind and may be conflicting when one tries to integrate them into a phase diagram description. The critical evaluation of these data has to start with the sample preparation, to validate the type of experiment with respect to the specific alloy and the conclusions made. Experience from the Higher Order Evaluation Program shows that occasionally a mere reinterpretation of experimental findings can establish consistency. Other criteria in critical evaluations are the conformity with the underlying thermodynamic principles and, for example, the compatibility of ternary data with the binary boundary systems when one of the three concentrations is approaching zero.

Profound critical evaluation together with continuous updating are the selection criteria for safe investment in phase diagram compilations.

6. *Listing of Major Phase Diagram Sources*

Table 1 lists the numbers of materials systems covered by the individual phase diagram compilations. These figures are distinctly different from the sometimes quoted "number of phase diagrams" where different versions of the same phase diagram, different scale conversions, etc., are often added up to give misleadingly large figures.

For ternary systems, where phase diagrams are three-dimensional constructs, it is assumed that all cuts and projections are presented which are available and necessary for the understanding of the system. Where this is not the case this article refers to "selected diagrams."

7. *Notes and Comments on Table 1*

7.1 *Hansen 1*

'Der Aufbau der Zweistofflegierungen'; editor: M. Hansen; publisher: Springer Verlag, 1936.

This is the first compilation of the binary alloy systems, covering 828 systems.

7.2 *Hansen 2*

'Der Aufbau der Zweistofflegierungen'; editors: M. Hansen, K. Anderko; publisher: McGraw-Hill, 1958.

This book is a complete revision of Hansen 1. It contains 1334 systems, covering the literature to 1955–57. Compared to Hansen 1, the detailed critical evaluations are condensed to short presentations of the accepted data. Atomic percent presentation of diagrams is used. Later compilations (Massalski, Okamoto) still contain a considerable number of phase diagrams from the pioneering works of Hansen.

7.3 *Elliott*

'Constitution of Binary Alloys,' First Supplement; editor: R. P. Elliott; publisher: McGraw-Hill, 1965.

This book was published as a supplement to Hansen 2. It contains 1719 systems, covering the literature to 1961.

7.4 *Shunk*

'Constitution of Binary Alloys,' Second Supplement; editor: F. A. Shunk; publisher: McGraw-Hill, 1969.

This was published as the second supplement to Hansen 2 after Elliott. It is comprehensive for the period 1962–64 and contains 2380 alloy systems.

7.5 *Massalski*

'Binary Alloy Phase Diagrams'; editor-in-chief: T. Massalski; publisher: ASM International, 1986 (1st edition), 1990 (2nd edition).

Table 1
Compilations of phase diagrams and related data (arranged in increasing order of elements and year of publication).

Author/ editor	Systems covered	No. of systems covered	Critically evaluated	Period covered by bibliography	Description	Available format	Publication year	Comments
Hansen 1, 2	Binaries	1334	All	To 1957	Text, diagrams	Printed	1958	2nd edition
Elliot	Binaries	1719	All	To 1961	Text, diagrams	Printed	1965	Supplement to Hansen
Shunk	Binaries	2380	All	1962–64	Text, diagrams	Printed	1969	Supplement to Elliot
Moffat	Binaries	1920	Some	To 1999	Text, diagrams	Printed	1977 and ongoing	Loose-leaf
Massalski	Binaries	2965	Most	1900–91; to 1994 on CD-ROM	Text, tables, diagrams	Printed and CD-ROM	1990	2nd edition
Lyakishev	Binaries	>2276	All	To 1992	Text, tables, diagrams	Printed	1996–2000	In Russian
Okamoto	Binaries	2335	Most	1900–2000	Text, tables, diagrams	Printed	2000	Update of Massalski
Levinsky	Binaries	188	All	To 1994	Text, tables, diagrams	Printed	1997	p–T–X diagrams
Effenberg (‘Ternary Alloys’)	Ternaries	3500	All	1900–2000	Text, tables, diagrams	Printed	1988 and ongoing	Vols. 1–18

Villars, Prince, Okamoto	Ternaries	7380	No	1900–89	Tables, diagrams	Printed and CD-ROM	1995	Selected diagrams
‘Equilibrium Phase Diagrams’	2-, 3-, 4-, … component		Some	1900(?) and ongoing	Text, diagrams	Printed and CD-ROM	1964 and ongoing	Formerly ‘Phase Diagrams for Ceramicists,’ 12 volumes
Effenberg (‘Red Book’)	1-, 2-, 3-, 4-, … component	All systems published	No	1990 and ongoing	Text, tables, diagrams	Printed	1990 and ongoing	18 books published
MSIT Workplace (an interactive working environment)								
Literature evaluations	Ternaries	3600	All	1900 and ongoing	Text, tables, diagrams	Electronic	1996	Also contains Effenberg ‘Ternary Alloys’
Literature contents	1-, 2-, 3-, 4-, … component	8000	No	1990 and ongoing	Text, tables, diagrams	Electronic	1996	Also contains Effenberg ‘Red Book’
Literature references	1-, 2-, 3-,4-, … component	100 000	NA	1900 and ongoing	Bibliography	Electronic	1996	
BRB	Binaries	1500	Partially	1990 and ongoing	Tables, diagrams	Electronic	2001	Not yet public
P–T–x	Binaries	188	All	To 1994	Text, tables, diagrams	Electronic	2000	Identical to Levinsky
Science forum	Internet-based “global workshops,” information channels such as “ongoing work” and “discussion groups”					Electronic	2000	

The second edition contains 2965 binary systems, most of which were critically evaluated within the International Binary Evaluation Program; 450 systems are adapted from earlier work. This compilation is the accepted standard reference work. It describes the phase relationships in tables and diagrams, both with accompanying text.

7.6 Moffatt

'Binary Phase Diagrams Handbook'; editor: W. G. Moffatt; publisher: General Electric, 1977 and ongoing.

This handbook includes systems that are not gathered in the compilations listed above, according to the editor. In addition to actual phase diagram information, this compilation includes an index that cross-references binary systems found in Moffat and the other compilations on binary systems. Updates of this handbook have been published biannually since 1977 as loose-leaf collections to replace or to supplement the preceding data sheets.

7.7 Lyakishev

'Phase Diagrams of Binary Metallic Systems'; editor-in-chief: N. P. Lyakishev; publisher: Mashinostroenie, 1996–2000 (in Russian).

The compilation consists of four volumes presenting over 2276 systems. In the critical evaluations, the authors use both original articles and earlier compilations. This work differs from the other binary compilations in including lattice parameters.

7.8 Okamoto

'Desk Handbook: Phase Diagrams for Binary Alloys'; editor: H. Okamoto; publisher ASM International, 2000.

This compilation is a revised version of Massalski to which the present editor contributed substantially. It provides updates for about 600 systems and 170 new systems. It aims to select the "best" diagram for each system and usually gives one or two key references per system.

7.9 Levinsky

'Pressure Dependent Phase Diagrams of Binary Alloys'; author: Yu. Levinsky; editor: G. Effenberg; publishers: MSI, ASM International, 1997.

The p–T–x compilation by Yu. Levinsky is based on the evaluation of over 12000 individual publications. A total of 188 binary systems is reviewed. Data are presented for p–T, p_{part}–T, and p_{part}–p diagrams, and their isobaric and isothermal sections. The data for transformations in the condensed state are described concisely and only where they differ from or are lacking in other compilations.

7.10 Effenberg

'Ternary Alloys'; editor-in-chief: G. Effenberg; publisher: MSI, 1988 and ongoing.

This series makes up part of the MSIT phase diagram program, carried out since 1984. All "system reports" are thoroughly evaluated and uniformly structured. The series proceeds by metal–X–Y categories, within a category in alphabetical order. There are 18 volumes available covering the Me–X–Y categories with Me = Ag, Al, Au, Li, Mg. At the time of publishing a category contains all ternary system for which data are available. It is the only series that provides critical evaluations for all systems, as far as published data allow.

7.11 Villars, Prince, Okamoto

'Handbook on Ternary Alloy Phase Diagrams'; editors: P. Villars, A. Prince, H. Okamoto; publisher: ASM International, 1995.

This handbook contains data on 7380 ternary systems. The bibliography covers literature until 1989. Selected "best choice" phase diagrams are presented in this 10-volume compilation together with a very short text, but there are 43695 references and extensive crystallographic data. The references have an identifier that allows the full paper to be ordered from ASM International.

7.12 'Phase Equilibria Diagrams'

Formerly 'Phase Diagrams for Ceramists'; editors: different over 12 volumes; publishers: AcerS and NIST, 1964–96 and ongoing.

This series is a joint effort of the American Ceramic Society and the National Institute of Standards and Technology, USA. It presents ceramic systems and started at the time that Hansen worked on the metal systems. It is a collection of 12 hardcover volumes (1964–96), five softbound issues (1991–97), a zirconium monograph (1998), and a CD-ROM (1997), covering oxides, salts, high-pressure systems, semiconductors, chalcogenides, borides, carbides, and nitrides. The number of published phase diagrams is given as about 25000.

7.13 Effenberg

'Red Book, Constitutional Data and Phase Diagrams, Annual Summaries of the World Literature'; editor-in-chief: G. Effenberg; publisher: MSI, 1993.

This book series extracts annually from over 300 journals and symposia information related to phase diagrams, thermodynamics, and crystallography, and presents about 120 previously unpublished papers per year. Covered are two-, three-, four-, five- (and more) component systems, whatever is published, starting with the publication year 1990.

The information is arranged by element combinations. Each summary extracts from each paper those and only those data that are relevant for the respective element combination. The summaries are translated into English, uniformly structured, and consist of text, tables, diagrams, and reference to the original sources. Materials property data are included where available within the above scope. A total of 18 books, over 10000 pages, has been issued.

The series is integrated in the MSIT Workplace and provides virtually full control over all publications on phase diagrams.

7.14 MSIT Workplace

The Materials Chemistry Knowledge Network. This is a working environment consisting of databases, communication platforms, and a global team of scientists, addressing phase diagrams and related data for materials systems, including topical workshops.

The MSIT Workplace identifies all element combinations that have ever been studied, makes this information available, and maintains the information up to date, on different levels:

(i) Literature citations: since 1900 and ongoing, over 100000 links from element combination to the literature.
(ii) Literature contents: all systems since 1990 and ongoing; see ‘Red Book’.
(iii) Literature evaluations: ternary systems; see ‘Ternary Alloys’ from MSI.
(iv) Ongoing work and moderated discussions.

The MSIT Workplace is the outlet of the largest phase diagram program of all, on which over 100 scientists from four continents have collaborated continuously since the mid-1980s.

This interactive system integrates and exceeds the ‘Ternary Alloys’ (MSI) and ‘Red Book’ series, which both originate from the same team. For more information see http://www.msiwp.com/

G. Effenberg
Materials Science International Services GmbH, Stuttgart, Germany

Polymer Crystallization: General Concepts of Theory and Experiments

The transformation of an amorphous crystallizable polymer into a semicrystalline material is not instantaneous even under favorable thermodynamic conditions. The subject of polymer crystallization is concerned with the description of this process, both from the phenomenological and the molecular viewpoint. The properties of semicrystalline polymers depend markedly upon the kind and distribution of the nanocrystals, as well as upon the total degree of crystallinity (Gedde 1995). Consequently, the control of physical properties is only possible through an understanding of the crystallization kinetics and the underlying molecular processes. Results of such studies shed light upon polymer crystallizability and morphology (Bassett 1981), which is directly related to kinetics.

Under appropriate conditions of temperature, pressure, and mechanical stress, or under the influence of solvents, a spontaneous three-dimensional arrangement of the molecules can take place (Bassett 1981). This orderly arrangement is called the *crystalline state*. Here the molecular organization is rather complex due to the fact that the crystallizing entities are not isolated structure elements, but interconnected by many molecular segments (tie molecules).

In a phase transformation such as crystallization, the two basic processes are the nucleation process by which a new phase is initiated within the parent phase, and the subsequent growth of the new phase (Wunderlich 1976). Nucleation and growth are concurrent processes that have been extensively analyzed in monomeric systems. The former may occur homogeneously, by statistical fluctuations in the parent phase, or by formation of nuclei catalyzed by heterogeneities or impurities present in the melt.

Figure 1 illustrates schematically a model of crystalline lamellae and its associated structural features: the crystalline region with a three-dimensional ordered structure, the interfacial zone, and the amorphous region consisting of chain segments in disordered conformations connecting adjacent crystallites. The crystal thickness l_c depends on crystallization temperature T_c and consequently a wide range of l_c can be obtained. The average crystal thickness can be determined from small-angle x-ray scattering (SAXS) (Baltá-Calleja and Vonk 1989) or from the Raman low frequency longitudinal mode (Gedde 1995). The ribbon-like lamellae organize themselves into higher ordered structures showing spherical symmetry (spherulites), when crystallized from the melt. Spherulites are spherically symmetric birefringent structures that display a Maltese cross when viewed under crossed polarizers using an optical microscope. Other crystalline morphologies vary from rods and sheet-like aggregates to a random distribution of lamellae (Bassett 1981).

For most flexible polymers crystallization from the melt takes place well below T_m. Thus high undercooling is required for crystallization. The reason for this behavior is related to the high interfacial free energy associated with the basal planes of the laminar crystallites and the difficulty in extracting

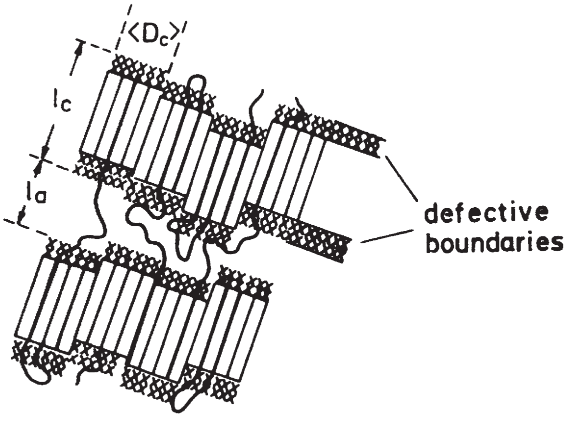


Figure 1
Schematic representation of lamellar nanostructure in semicrystalline polymers. The model shows lamellae of thickness l_c with interfacial defect boundaries alternated by amorphous regions of thickness l_a. D_c denotes the average coherent length of the crystals (from Baltá-Calleja *et al.* 2000).

ordered sequences of sufficient length from the disordered melt. Polymer crystallization may occur either from a deformed melt by application of an external force or from a quiescent melt. In the case of polymer glasses crystallization takes place in the temperature window between the glass transition temperature, T_g, and T_m (cold crystallization) (Gedde 1995).

Crystallization is accompanied by the evolution of latent heat (Wunderlich 1976). This property can be used to follow the crystallization kinetics by calorimetric methods. Traditional techniques use dilatometry (Mandelkern 1989). More recent methods include wide-angle x-ray scattering (WAXS) and SAXS using synchrotron radiation Wutz *et al.* (1995), Fourier transformed infrared spectroscopy and nuclear magnetic resonance (Gedde 1995), neutron scattering methods and dielectric spectroscopy (Nogales *et al.* 1999a). In the case of glassy polymers the microhardness technique has been recently used to follow the kinetics of cold crystallization *in situ* (Rueda *et al.* 1994, Baltá-Calleja and Fakirov 2000).

1. Density Fluctuations as Precursors of Crystallization

In isotropic polymers, crystallization usually proceeds through nucleation and growth mechanisms (Strobl 1996) leading to the formation of a micrometer scale organization of crystalline lamellae arranged within spherulites, axialites, or other supramolecular structures (Wunderlich 1976, Strobl 1996). However, in polymers, nucleation and growth is not the only mechanism inducing the building-up of a microphase-separated system consisting of an amorphous and crystalline phase. Oriented amorphous polymers, for example, may exhibit crystallization processes governed by the spinodal decomposition (SD) mechanism (Strobl 1996). Due to the fact that the formation of a nucleus implies the appearance of surfaces, the nucleation step requires an induction time where no crystals are present.

In the induction period, precrystallization processes may be present tending to promote a parallel arrangement of chain segments creating a density fluctuation across the system. This process, which is in agreement with pioneering ideas of Flory, starts with a local parallel organization of polymer chain segments, due to intramolecular interactions, followed by a subsequent increment in the intermolecular interaction between them. Recently, simultaneous WAXS and SAXS experiments performed in isotropic melts during the induction time of crystallization have revealed the existence of an excess of scattering at low angles. The latter has been interpreted as a signature of ordering phenomena acting as precursors for the crystallization process (Imai *et al.* 1992a, 1992b).

These experiments included initially semiflexible polymers (Imai *et al.* 1993, Ezquerra *et al.* 1996), but lately the range of polymers exhibiting this effect has been broadened to include flexible polymers Terrill *et al.* (1998), very rigid polymers Ania *et al.* (1999) and low molecular weight systems (Semmelhack and Esquinazi 1998). To illustrate these phenomena, Fig. 2 shows the simultaneous WAXS and SAXS patterns taken during an isothermal cold crystallization experiment ($T_c = 270\,°C$) on new thermoplastic poly-imide (TPI) Ania *et al.* (1999). Here, it is clearly seen that in the early stages of crystallization, an excess of scattering at low angles (SAXS) rendering the formation of a maximum around $q = 0.5\,nm^{-1}$ emerges before the appearance of Bragg maxima in the WAXS patterns.

Figure 3 shows the time evolution of the normalized integrated SAXS intensity as a function of time for three different crystallization temperatures (265 °C, 270 °C, and 272 °C). It is seen that the delay time between the appearance of SAXS scattering (indicated by upward arrows) and the onset of crystallinity (denoted by downward arrows) increases as crystallization temperature decreases. The time evolution of the SAXS excess of scattering appears to follow the kinetics of a SD mechanism as described by the Cahn-Hilliard theory developed for metal alloys. For homopolymers, a phenomenological theory has been developed in which it is shown that the coupling between density and chain conformation may induced a liquid–liquid bimodal within the equilibrium liquid-crystalline solid coexistence region Olmsted *et al.* (1998).

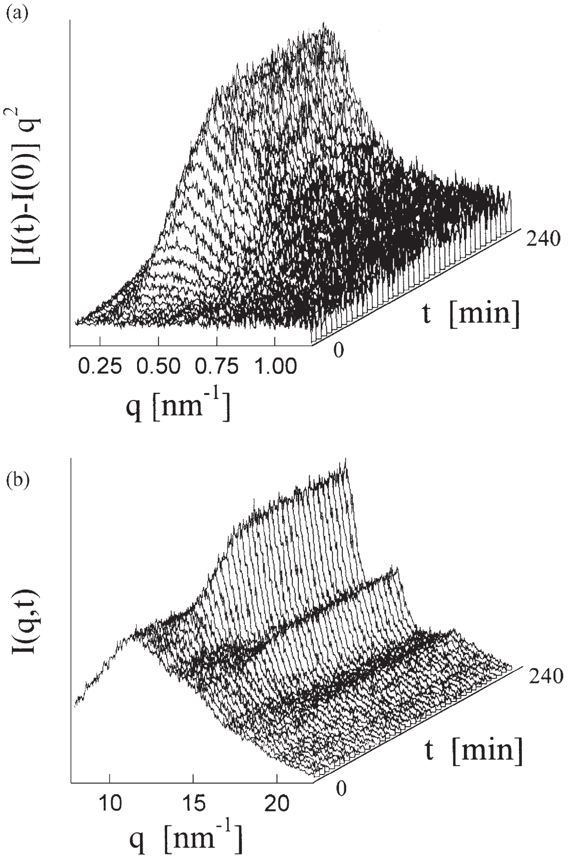


Figure 2
Variation of (a) small- and (b) wide-angle x-ray scattering during isothermal treatment of New-TPI at $T_c = 270\,°C$ from the glass. The patterns were obtained simultaneously using synchrotron radiation.

2. *Bulk Crystallization*

2.1 Spherulite Growth

The kinetics of crystallization is frequently analyzed by polarizing microscopy. The spherulites are usually circular in shape during the initial stages of crystallization, and the radii are measured before they impinge on their neighbors (Gedde 1995, Bassett 1981, Wunderlich 1976). The temperature range over which the growth rate can be measured depends on the number and size of the spherulites and the latter depend, in turn, on the nucleation and growth rates. The growth rate may be so high that the dissipation of the heat of crystallization prevents the attainment of isothermal conditions. In some cases the light scattering technique is used to derive the spherulite radius also for measurements of the growth rate (Gedde 1995).

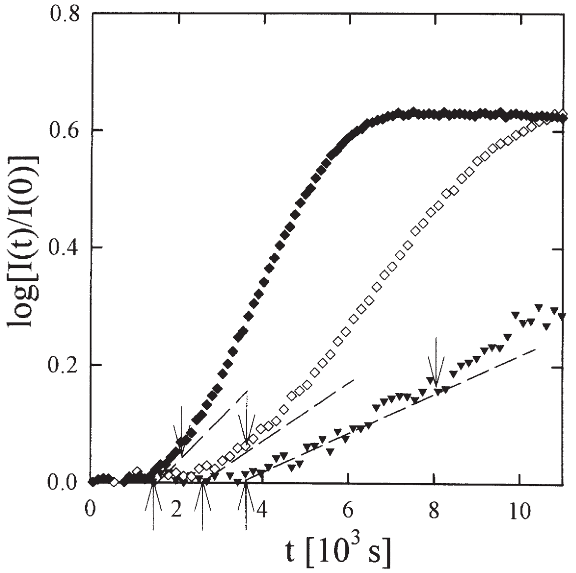


Figure 3
Variation of the integral intensity, $\log[I(t)/I(0)]$, in the range 0.12–1.15 nm^{-1} as a function of time for various temperatures: $T_c = 265$ (▼), 270 (◇), and 272 °C (◆). Downward arrows denote the onset of crystallization. Upward arrows indicate the appearance of SAXS scattering.

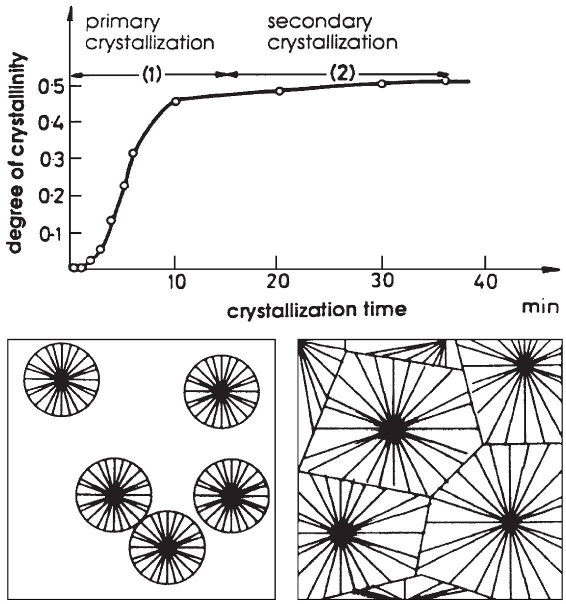
2.2 Crystalline Phase Evolution: Avrami Analysis

In the course of isothermal crystallization of polymers one observes a sigmoidal rapid increase of the degree of crystallinity x_c followed by a very slow increase (Zachmann and Stuart 1960) (Fig. 4 top). The comparatively rapid first increase is called *primary crystallization* while the following much slower process is denoted as *secondary crystallization*. In the polarizing microscope one observes that during primary crystallization spherulites are growing outward from nuclei until impingement produces polygonal structures and termination of the radial growth results (Fig. 4 bottom). During primary crystallization to a first approximation one can assume that the degree of crystallinity within the spherulites is constant.

The sigmoidal form of the curve can then be explained by applying the theory of Avrami to the growth of the spherulites. In this theory one takes into account the impingement of morphological units like spherulites with statistically distributed centers. One then obtains:

$$X_s = 1 - \exp(-kt^n)$$

which relates X_s the amount or fraction of molecules transformed into spherulites or other morphological



(a) (b)

Figure 4
(top) Typical variation of degree of crystallinity as a function of time during isothermal crystallization of polymers (from Zachmann and Stuart 1960); (bottom) schematic representation of structural changes (spherulite growth) during primary and secondary crystallization.

units after some time t, at a given crystallization temperature, to the growth rate parameter k and mode of nucleation n. The overall transformation rate k embodies both nucleation and growth (presumed by Avrami to occur under isovolume conditions). The Avrami approach was later modified by Mandelkern to include incomplete crystallization (Mandelkern 1989). The analysis of the overall crystallization data for many polymers follows quite well the above equation for the initial crystallization process.

The value of n can be obtained from the slope of a double logarithmic plot, $\log \cdot \log X_s$ against $\log t$. Integral values between $n = 1$ and $n = 4$ have been found in many polymers. Usually it is assumed that n is 1, 2, or 3 according to whether crystal growth takes place in one, two, or three directions at once (rods, disks, spheres, respectively) (Wunderlich 1976). However, in other cases fractional data fitting have been reported. For $n = 4$, growth is spherulitic from embryos sporadically nucleated in time and position.

During secondary crystallization X_s stays constant equal to 1 while the crystallinity x_c slowly increases (Zachmann and Stuart 1960, Zachmann and Wutz 1993). There are three possible mechanisms to explain such an increase: (i) the crystallites may become thicker while the amorphous regions become thinner; (ii) new crystals may grow within the lamellar stacks; or (iii) in the case that the spherulites are not completely filled by lamellar stacks, new lamellar stacks consisting of lamellar crystals separated by noncrystalline regions may be formed. By means of simultaneous measurements of WAXS, SAXS, and light scattering (LS), it is possible to distinguish between the three possible processes to explain secondary crystallization. From LS measurements one can derive the radius of the spherulites R as a function of time and thus determine the end of primary crystallization where R becomes constant. From the change of the crystal reflections intensity measured by WAXS one can derive the variation in the overall degree of crystallinity x_c. Finally, from the SAXS measurements, one can derive the change of the scattering power or invariants (Miller 1979) (scattering intensity integrated over all scattering angles). By using all this information in the evaluation of the results one can determine the mechanism of secondary crystallization (Zachmann and Wutz 1993).

2.3 Amorphous Phase Evolution

As described in the previous paragraph, the experimental input in crystallization comes to a large extent from x-ray scattering techniques. These techniques (SAXS and WAXS) provide information about the structure of the ordered regions at different length scales. Processes occurring in the amorphous fraction of the material are, however, "invisible" for these techniques due to the absence of order.

In glass-forming polymers, at temperatures above the glass transition temperature, T_g, there appears a characteristic relaxation process, called α-relaxation, which involves motions extended to several molecular segments. In semicrystalline polymers at $T > T_g$, some of the long amorphous polymeric chains are restricted to move between the crystalline regions. The occurrence of crystallinity in a polymer is revealed in the dynamics of the α-relaxation (in contrast to that of the pure amorphous polymer) by three specific effects: (i) a decrease of the intensity of the relaxation; (ii) an increase of the characteristic relaxation time; and (iii) a concurrent change in its shape (Ezquerra *et al.* 1994a, Dobbertin *et al.* 1996). In particular, dielectric spectroscopy experiments (DS) have shown that, upon crystallization, the dynamics of the α-relaxation is strongly affected by the progressive development of the crystalline phase (Ezquerra *et al.* 1994a, 1994b).

Figure 5 shows the real-time evolution of the α-relaxation during an isothermal crystallization process at $T_c = 188\,^\circ\mathrm{C}$ for poly(ether ketone ketone) (PEKK) (Ezquerra 1994b). Figure 5 (top) and (bottom) show the dielectric loss, ε'', and dielectric constant, ε', values respectively. In Fig. 5 (top) every curve presents the

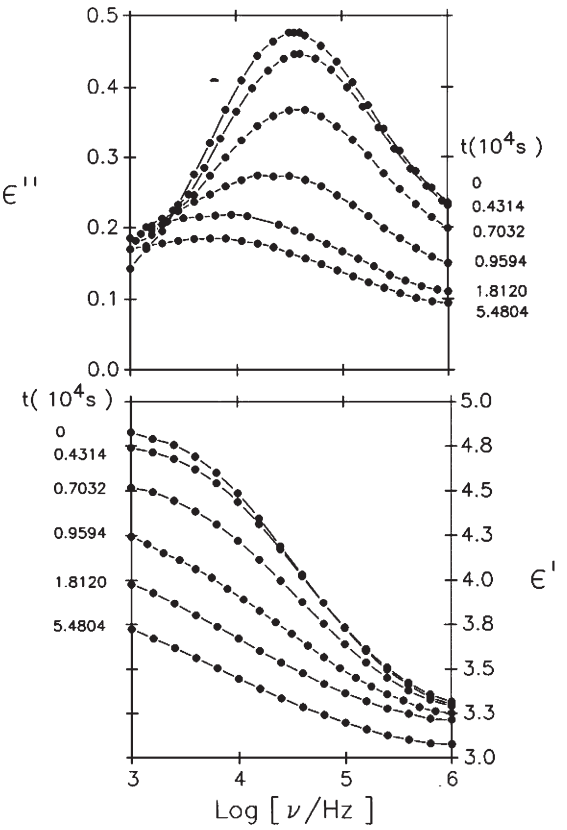


Figure 5
Real-time evolution of the α-relaxation dielectric loss, ε″ (top), and dielectric constant, ε′ (bottom), values versus frequency during an isothermal crystallization process at $T = 188\,^{\circ}\text{C}$ at selected crystallization times.

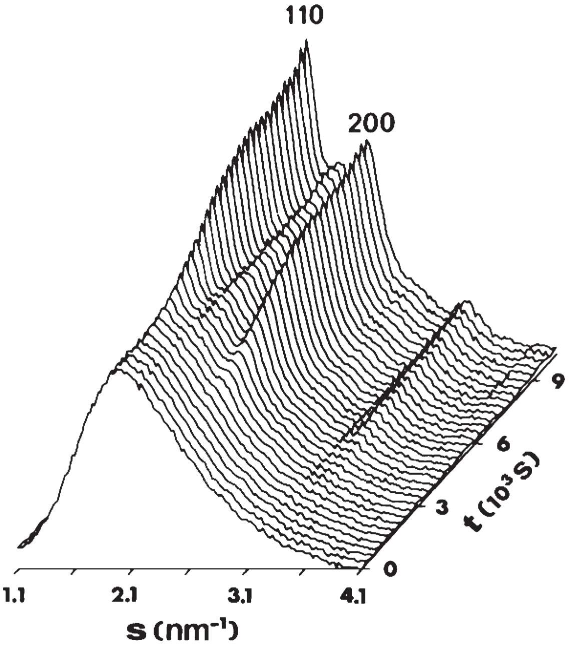


Figure 6
Three-dimensional plot showing the real-time evolution of the WAXS patterns with crystallization time (t_c) at $T_c = 188\,^{\circ}\text{C}$ of PEKK. The diffracted x-ray intensity has been represented as a function of the reciprocal lattice vector $s = (2\sin\theta)/\lambda$ for each time in seconds.

dependence of ε″ versus frequency at a selected crystallization time (labeled on the right side of the figure). The initial curve, at $t = 0$, corresponds to the initially amorphous polymer. Figure 6 shows the variation of the WAXS pattern in real time during isothermal crystallization of PEKK at $T_c = 188\,^{\circ}\text{C}$. The x-ray scattered intensity is represented as a function of the reciprocal lattice vector $s = (2\sin\theta)/\lambda$ where 2θ is the scattering angle. As crystallization time increases the 110 ($s = 2.09\,\text{nm}^{-1}$) and 200 ($s = 2.58\,\text{nm}^{-1}$) reflections, characteristic of an orthorhombic unit cell, appear Gardner *et al.* (1992).

As crystallization proceeds three main features are observed in Fig. 5 (top). First, there is a reduction of the intensity of the α-relaxation process as the crystallization time increases. Second, a shift of the frequency of maximum loss towards lower frequency values is detected. Third, a concurrent change of the shape parameters is observed. Some of these features are illustrated in Fig. 7, which shows the plot of the logarithm of the maximum loss frequency, $\nu_{\max}$, as a function of the crystallization time (t_c). If one considers that $(2\pi\nu_{\max})^{-1}$ is an average relaxation time, the observed decrease of $\nu_{\max}$ suggests a slowing down of the overall chain mobility as crystallization occurs. Another important parameter that can be extracted from the dielectric experiments is dielectric strength, Δε, which is related to the area under the relaxation process. Since Δε is proportional to the density of dipoles involved in the α-relaxation, the decrease of Δε with t_c can be associated with the progressive reduction of the amorphous phase as segments of the polymeric chains crystallize. Figure 8 shows the correlation found between Δε and the x-ray crystallinity. The straight line drawn represents the best linear fit to the experimental data. As shown, the straight line extrapolates to $\Delta\varepsilon = 0$ at a crystallinity value different from 100%. This finding suggests that the restriction in the motion of polymeric chains in the amorphous regions due to adjacent crystalline regions is higher than expected only on the basis of the amorphous content. From this result one may infer the existence of a tightly bound amorphous phase of rigid disordered chains, which are constrained under the influence of the crystal lamellae (Nogales *et al.* 1999a, 1999b, Cheng *et al.* 1986).

3. Solution Crystallization

Crystallization from solution has been investigated in concentrated solutions, polymer–diluent systems, and diluted solutions. Concentrated solutions are characterized by the random distribution of polymer segments throughout the mixture, and in many systems the diluent is prevented from entering the crystal lattice. These differences are important in the crystallization mechanisms. There are few studies of the kinetics of concentrated solutions and dilute crystallization has been less extensively studied than bulk crystallization despite the numerous theoretical and experimental investigations of the resulting morphology (Gedde 1995).

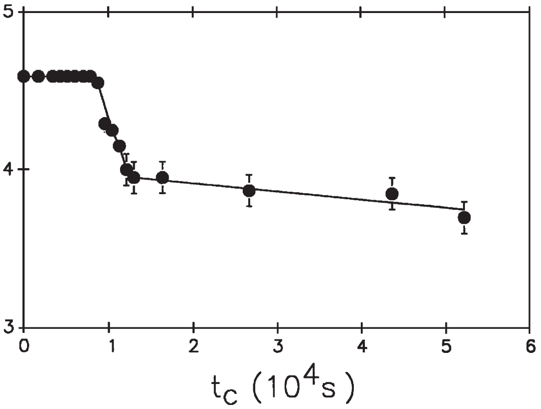


Figure 7
Variation of the frequency of maximum loss values ν_{max} corresponding to Fig. 5 (top) as a function of crystallization time t_c.

Following the discovery in 1957 that polymeric materials could crystallize as very thin lamellar single crystals (~10 nm thick and several μm in lateral dimensions) from dilute solution, it became evident that the "fringed micelle model" (Wunderlich 1976) was no longer consistent with the new knowledge gained at that time. The orientation of the chain axis is preferentially perpendicular to the wide face of the flat crystals. The latter thickens during growth through a screw dislocation mechanism leading to a spiral terrace structure (Wunderlich 1976). Since the average length of the molecules is much larger than the crystal thickness, the inescapable conclusion is that the lamellae present a chain-folded structure. Under stress-free conditions, chain folding appears to be the preferred mode of growth from dilute solution.

Very stiff rod-like synthetic molecules cannot fold, and therefore, form glasses or liquid crystals. Stiff polymer molecules usually contain aromatic rings, exhibiting good thermal properties, high glass temperatures, and few single bonds. Therefore these stiff-chain polymers show considerable inflexibility, insolubility, and intractability. Advanced polymers

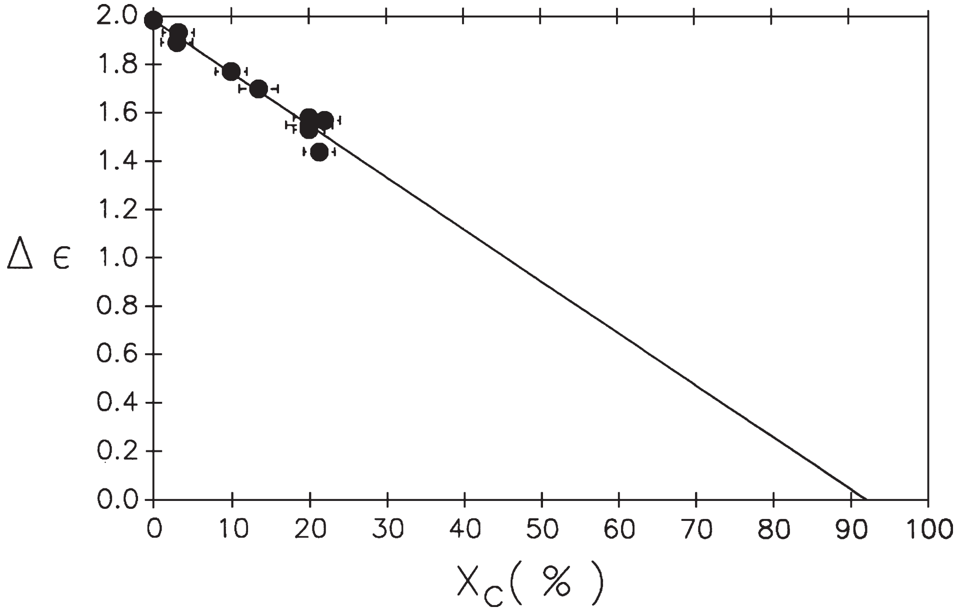


Figure 8
Variation of Δε with crystallinity as derived from WAXS experiments. The straight line represent the best linear dependence to the experimental data.

such as poly(ether ether ketone) (PEEK), PEKK, poly(ethylene naphthalate) (PEN), and new TPI, are examples of rigid backbone that belong to this class (Ezquerra *et al.* 1996, Nogales *et al.* 1999b). Several natural polymers (starch, etc.) also fall into this category, even though they can crystallize under appropriate conditions.

The thickness of solution-grown lamellar crystals tends to be uniform, for a given solvent, at a given crystallization temperature, and depends strongly on the crystallization temperature. However, the thickness is independent of molecular weight, except for low molecular weights Ungar *et al.* (1985). A great variety of crystal habits have been observed from lamellae to dendritic forms (Bassett 1981). The crystal habits depend on the crystallization temperature or undercooling, on the polymer–solvent interaction, and on the solution concentration (Keller 1986).

Crystallinity of solution-crystallized polymers is substantially higher than that of the corresponding bulk-crystallized samples. Although solution-formed crystals thicken as a function of time during annealing, thickening does not take place during isothermal crystallization from dilute solution (Organ and Keller 1985). This fact, which is quite different from the bulk crystallization behavior, can be explained by the high supercooling involved in the crystallization process of dilute systems.

4. Thermodynamics of Crystal Formation: Growth Theories

4.1 Chain-folded Crystals: Surface Free Energy

Nowadays it is accepted that kinetic factors control growth rate and morphology and equilibrium theories are no longer considered. Kinetic theories accept that the end state is not that with the lowest possible free energy. Growth rate depends on crystal thickness. Crystals with a range of crystal thicknesses larger than a minimum value $2\sigma_e/\Delta h_f$ are formed, where σ_e is the surface-fold free energy of the crystal and Δh_f is the enthalpy of fusion. The resulting distribution of crystal thickness values is determined by the relationship between crystallization rate and the crystal thickness at a given temperature. A maximum in growth rate is obtained at a crystal thickness value which is larger than $2\sigma_e/\Delta h_f$ by a small term δl_c. The Lauritzen–Hoffman theory assumes that the free energy barrier associated with nucleation has an energetic origin (Lauritzen and Hoffmann 1973). The Sadler–Gilmer theory regards the free energy barrier as predominantly entropic (Sadler and Gilmer 1984). Kinetic theories predict the temperature dependence of crystal growth rate, initial crystal thickness l_c^* and other morphological parameters. The theory of Lauritzen–Hoffman provides an expression for the linear growth rate (G) at which spherulites or axialites grow radially as a function of degree of supercooling ($\Delta T = T_m^\circ - T_c$) where T_m° is the equilibrium melting point and T_c is the crystallization temperature. A relatively large barrier must be overcome to initiate the nucleus, after which additions are accompanied by chain folding. This results in a net gain in free energy until a stable nucleus is formed. Stability is indicated by a maximum in the relationship between the Gibbs free energy of formation and the number of elements in the step.

If we denote *ab* the cross-section of the polymer chain, where *b* is the growth direction, then the chain spreads gradually across the substrate (Fig. 9). After this initial chain attachment, additional chains adhere themselves once more to this interface. These chains then fold across the surface, thereby extending the growth front and crystal dimensions at a rate that depends on the crystallization undercooling. The free energy of crystal formation can be written as:

$$\Delta G = 2bl\sigma + 2ab\sigma_e - abl(\Delta f)$$

σ being the lateral surface free energy of the crystal.

Regime I growth is characterized by the fact that the secondary nucleation step controls the linear growth rate (G). The lateral growth rate (g) is significantly larger than the rate of formation of secondary nuclei (i)

$$g \gg i$$

The whole substrate is completed and covered by a new monolayer. Monolayers are added one by one and the linear growth rate is given by

$$G_I = biL$$

where i is the surface nucleation rate (nuclei per length of substrate per second) and L is the substrate length, $L = n_s a$, and n_s is the number of the stems on the substrate.

Regime II growth takes place by multiple nucleation. The secondary nucleation rate is in this case more rapid than regime I, i.e., $g < i$. Regime III growth occurs by prolific multiple nucleation. The niche separation in this case is the same size as the stem width. It can be shown that the growth rate (G_{III}) is given by

$$G_{III} = biL$$

The Lauritzen and Hoffman theory compares favorably with experimental data in two important aspects: the temperature dependence of both the initial crystal thickness l_c^* and the linear growth rate.

The theory has been modified several times (Lauritzen and Hoffmann 1973, Hoffmann *et al.* 1979, Hoffmann and Miller 1997). Crystallization of multicomponent chains and fold-length fluctuations are phenomena which were addressed in different revisions during the 1960s. Diffusion plays an important role in crystallization, and the view of reptation was

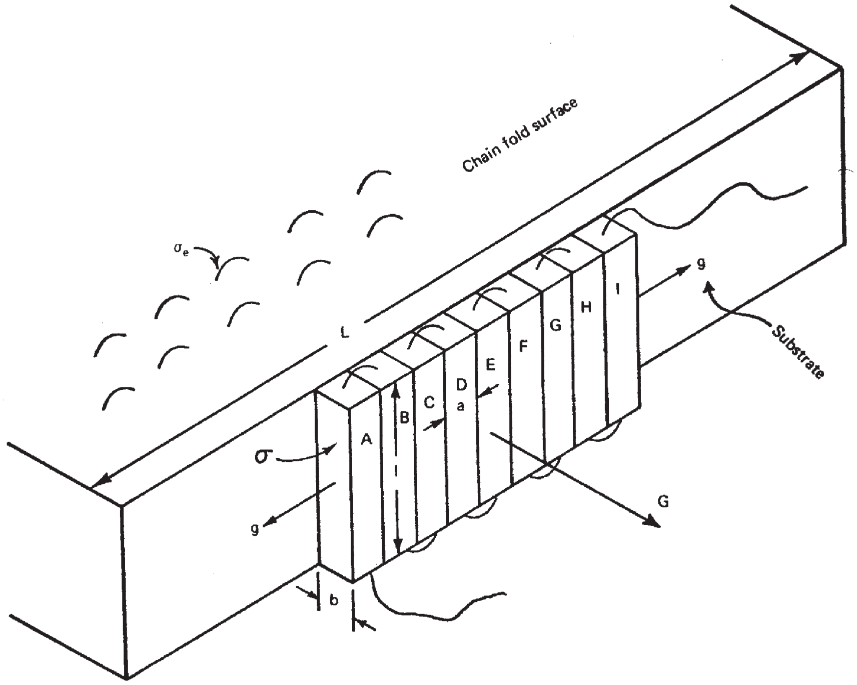


Figure 9
Model for surface growth of a chain-folded crystal on a substrate of width L and strip height l and width a. Growth occurs as chains nucleate and spread across the face. G is the growth rate and g is the spreading rate on the substrate.

implemented Hoffmann *et al.* (1979). Hoffman and co-workers showed that the self-diffusion rate was sufficiently high to ensure a sizeable proportion of tight chain folding. This development led to a modification of the growth rate equation. The linear growth rate in all three regimes was shown to be proportional to $1/M_w$. This expression was introduced into the pre-exponential factor of the growth-rate equations. The last modification addresses the problem of the curved edges of crystals grown from the melt at high temperature, clearly within the regime I temperature region. The elliptical shape of the crystal is microfaceted with steps of about 100 nm. The linear growth rate becomes approximately equal to the lateral growth rate (g).

Criticism of the Lauritzen–Hoffman theory has appeared (Point and Villars 1992). The theory predicts a growth-rate transition, essentially because g changes only moderately with temperature whereas i varies strongly. Point and Dosière (1989) suggest other causes for the change in the slopes accompanying the regimes I–II transition: molar mass segregation, temperature dependence of the interfacial energies, viscosity effects, and temperature dependence of the nucleation processes.

An updated study of the rate of growth of chain-folded lamellar crystals from the subcooled melt of polyethylene fractions has been treated by Hoffman and Miller in terms of surface nucleation theory aiming to illuminate the origin of chain-folding and associated kinetic effects in molecular terms (Hoffmann and Miller 1997). Key experimental data have been analyzed in terms of the theory and essential parameters determined, including the size of the substrate length L involved in regime I growth. New evidence based on the quantization effect indicates a high degree of adjacent re-entry in regime I for lower molecular weight fractions. The nature of the chain folding at higher molecular weights for the various regimes has also been discussed in terms of kinetic, neutron scattering, IR, and other evidence.

4.2 Extended-chain Crystals

Extended-chain crystals (ECCs) have been observed on crystallization of the melt under high pressure (Wunderlich 1976) and are formed from folded-chain crystals through lamellar thickening. In polyethylene the ECC structure arises when the melt crystallizes into the hexagonal phase and the chain-folded structure into the orthorhombic phase, as described by Bassett (1981). The lateral growth rate (V) of an isolated extended-chain single crystal (ECSC) has been observed by means of optical microscopy under isobaric conditions and the temperature dependence of

ECSCs led to the conclusion that the lateral growth of these crystals is nucleation controlled. Hikosaka has proposed a new kinetic theory called "chain sliding diffusion theory" (Hikosaka 1990). This theory shows that the origin of ECCs and chain-folded crystals is related to the respective ease and/or difficulty of chain sliding diffusion within the nucleus and/or crystals. The concept of sliding diffusion along the chain axis emphasizes the topological nature of long chain molecules. The theory explains the observed ΔT dependence of V for an ECSC of PE. It also shows that the lamellar thickening growth rate (U) is large when polymers crystallize into the mobile phase, where sliding diffusion is distinguishable, and gives rise to the formation of an ECSC. However, U is small when polymers crystallize into the immobile phase, such as the orthorhombic phase, where sliding diffusion is slow, giving rise to chain-folded crystal formation. In summary, the theory predicts that polymers which crystallize from the melt into the mobile phase form ECCs, while polymers crystallizing into the more ordered phase will form chain-folded crystals. These predictions have been confirmed on several polymers.

Crystallization experiments of an ECSC of PE under high pressure from the melt into the mobile phase have provided clear evidence that an isolated ECSC is formed from a folded-chain single crystal through the combination of "lamellar thickening growth" and the traditional lateral growth concept. This finding reveals that lamellar thickness increases linearly with time (t); i.e., that the lamellar thickening growth rate is constant as long as the crystal is isolated and does not impinge with others. This result is different from the well-known lamellar thickening behavior of stacked-lamellae where l is known to increase linearly with $\log t$.

Bibliography

Ania F, Cagiao M E, Baltá-Calleja F J 1999 Density fluctuations as precursors of crystallization in a thermoplastic polyimide. *Polymer J.* **31**, 735–8

Baltá-Calleja F J, Fakirov S 2000 *Microhardness of Polymers.* Cambridge University Press, Cambridge

Baltá-Calleja F J, Vonk C G 1989 *X-ray Scattering of Synthetic Polymers.* Elsevier, Amsterdam

Bassett D C 1981 *Principles of Polymer Morphology.* Cambridge University Press, Cambridge

Cheng S Z D, Cao M Y, Wunderlich B 1986 Glass transition and melting behavior of poly(oxy-1,4-phenyleneoxy-1,4-phenylenecarbonyl-1,4-phenylene) (PEEK). *Macromolecules* **19**, 1868–76

Dobbertin J, Hensel A, Schick C 1996 Dielectric spectroscopy and calorimetry in the glass transition region of semi-crystalline poly(ethylene terephthalate). *J. Thermal Anal.* **47**, 1027–40

Ezquerra T A, Baltá-Calleja F J, Zachmann H G 1994a Real time dielectric relaxation of poly(ethylene terephthalate). *Polymer* **35**, 2600–6

Ezquerra T A, López-Cabarcos E, Hsiao B S, Baltá-Calleja F J 1996 Precursors of crystallization via density fluctuations in stiff-chain polymers. *Phys. Rev. E* **54**, 989–92

Ezquerra T A, Majszczyk J, Baltá-Calleja F J, López-Cabarcos E, Gardner K H, Hsiao B S 1994bb Molecular dynamics of the α-relaxation during crystallization of a glassy polymer: a real time dielectric spectroscopy study. *Phys. Rev. B* **50**, 6023–31

Gardner K H, Hsiao B S, Matheson R R Jr, Wood B A 1992 Structure, crystallization and morphology of poly(aryl ether ketone ketone). *Polymer* **33**, 2483–95

Gedde U W 1995 *Polymer Physics.* Chapman and Hall, London

Hikosaka M 1990 Unified theory of nucleation of folded-chain crystals (FFCs) and extended-chain crystals (ECCs) of linear-chain polymers: 2. Origin of FCC and ECC. *Polymer* **31**, 458–68

Hoffman J D, Guttman C M, DiMarzio E A 1979 On the problem of crystallization of polymers from the melt with chain folding. *Disc. Faraday Soc.* **68**, 177–97

Hoffman J D, Miller R L 1989 Response to criticism of nucleation theory as applied to crystallization of lamellar polymers. *Macromolecules* **22**, 3502–5

Hoffman J D, Miller R L 1997 Kinetics of crystallization from the melt and chain folding in polyethylene fractions revisited: theory and experiment. *Polymer* **38**, 3151–212

Imai M, Kaji K, Kanaya T 1993 Orientation fluctuation of poly(ethylene terephthalate) during the induction period of crystallization. *Phys. Rev. Lett.* **71**, 4162–5

Imai M, Mori T, Mizukami K, Kaji K, Kanaya T 1992a Structural formation of poly(ethylene terephthalate) during the induction period of crystallization: 1. Ordered structure appearing before crystal nucleation. *Polymer* **33**, 4451–6

Imai M, Mori T, Mizukami K, Kaji K, Kanaya T 1992b Structural formation of poly(ethylene tertephthalate) during the induction period of crystallization: 2. Kinetic analysis bssed on the theories of phase separation. *Polymer* **33**, 4457–62

Keller A 1986 Recent developments in morphology of crystalline polymers. In: Sedlácek B (ed.) *Morphology of Polymers.* Walter de Gruyter, Berlin, pp. 3–27

Lauritzen J I Jr, Hoffman J D 1973 Extension of theory of growth of chain-folded polymer crystals to large undercoolings. *J. Appl. Phys.* **44**, 4340–52

Mandelkern L 1964 *Crystallization of Polymers.* McGraw Hill, New York

Miller R L 1979 *Flow-induced Crystallization in Polymer Systems.* Gordon and Breach, New York

Nogales A, Ezquerra T A, Batallán F, Frick B, López-Cabarcos E, Baltá-Calleja F J 1999a Restricted dynamics in poly(ether-ether-ketone) as revealed by incoherent quasielastic neutron scattering and broad band dielectric spectroscopy. *Macromolecules* **32**, 2301–8

Nogales A, Ezquerra T A, García J M, Baltá-Calleja F J 1999b Structure-dynamics relationships of α-relaxation in flexible copolyesters during crystallization as revealed by real time methods. *J. Polym. Sci. B: Polym. Phys.* **37**, 37–49

Olmsted P D, Poon W C, McLeish T C B, Terril N J, Ryan A J 1998 Spinodal-assisted crystallization in polymer melts. *Phys. Rev. Lett.* **81**, 373–6

Organ S, Keller A 1985 Solution crystalization of polyethylene at high temperatures, part 3. The fold lengths. *J. Mater. Sci.* **20**, 1602–15

Point J J, Dosière M 1989 Crystal growth rate as a function of molecular weight in polyethylene crystallized from the melt:

an evaluation of the kinetic theory of polymer crystallization. *Polymer* **30**, 2292–6

Point J J, Villars D 1992 Crystals with curved edges: a unified model that mediates between the theories of nucleation-controlled and rough surface growth. *Polymer* **33**, 2263–72

Rueda D R, Viksne A, Malers L, Baltá-Calleja F J, Zachmann H G 1994 Influence of morphology on the microhardness of poly(ethylene naphthalene-2,6-dicarboxylate). *Macromol. Chem. Phys.* **195**, 3869–76

Sadler D M, Gilmer G H 1984 A model for chain folding in polymer crystals: rough growth faces are consistent with the observed growth rates. *Polymer* **25**, 1446–52

Semmelhack H C, Esquinazi P 1998 Observations of metastable ordered structures and kinetics of crystallization of o-terphenyl. *Physica B* **254**, 14–20

Strobl G 1996 *The Physics of Polymers*. Springer, Berlin

Terrill N J, Fairclough P A, Towns-Andrews E, Komanschek B U, Young R J, Ryan A J 1998 Density flucturations: the nucleation event in isotactic polypropylene crystallization. *Polymer* **39**, 2381–5

Ungar G, Stejny J, Keller A, Bidd I, Whiting M C 1985 The crystallization of ultra long normal paraffins: the onset of chain folding. *Science* **229**, 386–9

Wunderlich B 1976 *Macromolecular Physics. Crystal Nucleation, Growth, Annealing*. Academic Press, New York Vol. 2

Wutz C, Bark M, Cronauer J, Döhrmann R, Zachmann H G 1995 Simultaneous measurements of small angle X-ray scattering, wide angle X-ray scattering and light scattering during phase transitions in polymers. *Rev. Sci. Instrum.* **66**, 1303–7

Zachmann H G, Stuart H A 1960 Schmelz- und Kristallisationserscheinungen bei makromolekularen Substanzen. IV. Haupt- und Nachkristallisation von Terylen aus dem Glaszustand. *Makromol. Chemie* **49**, 131–47

Zachmann H G, Wutz C 1993 Studies of the mechanism of crystallization by means of WAXS and SAXS employing synchrotron radiation. In: Dosière M (ed.) *Crystallization of Polymers*. Kluwer, Dordrecht, The Netherlands, pp. 403–14

F. J. Baltá-Calleja and T. A. Ezquerra
Instituto de Estructura de la Materia, CSIC, Madrid, Spain

Polymer Spherulites

Spherulites and spherulitic growth are characteristic of melt-crystallized polymers and exert a profound influence on their mechanical and other properties. Recent work has shown that spherulitic growth in polymers, for a long time poorly understood, is a consequence of the molecular length, as well as being the cause of an inherent systematic spatial variation in texture-dependent properties.

Though widely encountered in nature, commonly forming from viscous and impure melts (Phillips 1994, Keller 1958, Keith and Padden 1963), spherulites are best characterized in crystalline polymers (Bassett 1984, 1994, Vaughan and Bassett 1989). A spherulite is literally a little sphere the shape of an isolated mature object, with a typical diameter of microns or more. However, the use of the term also needs to accommodate the variety of earlier forms actually observed. Specimens of isolated mature objects viewed in the polarizing microscope are circular, polycrystalline, and birefringent with a radial texture usually at or about the limit of optical resolution (Fig. 1a). A typical development is from an individual, apparently linear entity, such as a lamella viewed sideways, which branches repetitively and passes through sheaf-like forms in projection before attaining a near circular envelope. Large sheaf-like forms developed especially at higher temperatures, with reduced branching, are commonly known as hedrites (referring to their polyhedral habit in one perspective) or axialites because, to a first approximation, they consist of lamellae splaying around a common axis. All of these immature habits are, nevertheless, the result of spherulitic growth albeit with differing details, so that it is the growth process itself rather than the particular object which is the common feature and to which attention needs to be directed. This is the process whereby an object nucleated at a point branches, and the branches diverge, potentially leading, after sufficient iteration, to a spherical object of reasonably constant density with crystallographically equivalent radii.

The reference to point nucleation is necessary because in other circumstances there are different consequences. These include *transcrystalline layers* when growth is nucleated on, and proceeds perpendicular to a surface, as for i-polypropylene on poly(tetrafluoroethylene), and *row structures* growing radially out from linear nuclei (Fig. 1b). The two are different in that in the former only the growth direction is common, whereas in the latter the chain axis is common, being parallel to the nucleus with the growth directions radial. Despite the three different superstructures, which result from the different constraints placed on interlamellar development, all three geometries have the same isothermal radial growth rate (IRG) in homopolymers, namely that of the advance of an isolated individual lamella. The examination of row structures has been particularly informative in resolving outstanding issues of spherulitic growth because they have uniquely allowed the controlled observation and measurement of lamellar growth from its outset, in all perspectives.

1. *The Nature of Spherulitic Growth*

The two principal features of spherulites are the constant IRG, found in a very wide range of polymers (Keith and Padden 1963), although not universally, and the radial texture. Plotted against temperature over a sufficiently wide range, the former passes through a maximum reflecting the interplay between thermodynamic factors, which dominate nearer the melting point, and molecular transport, which slows

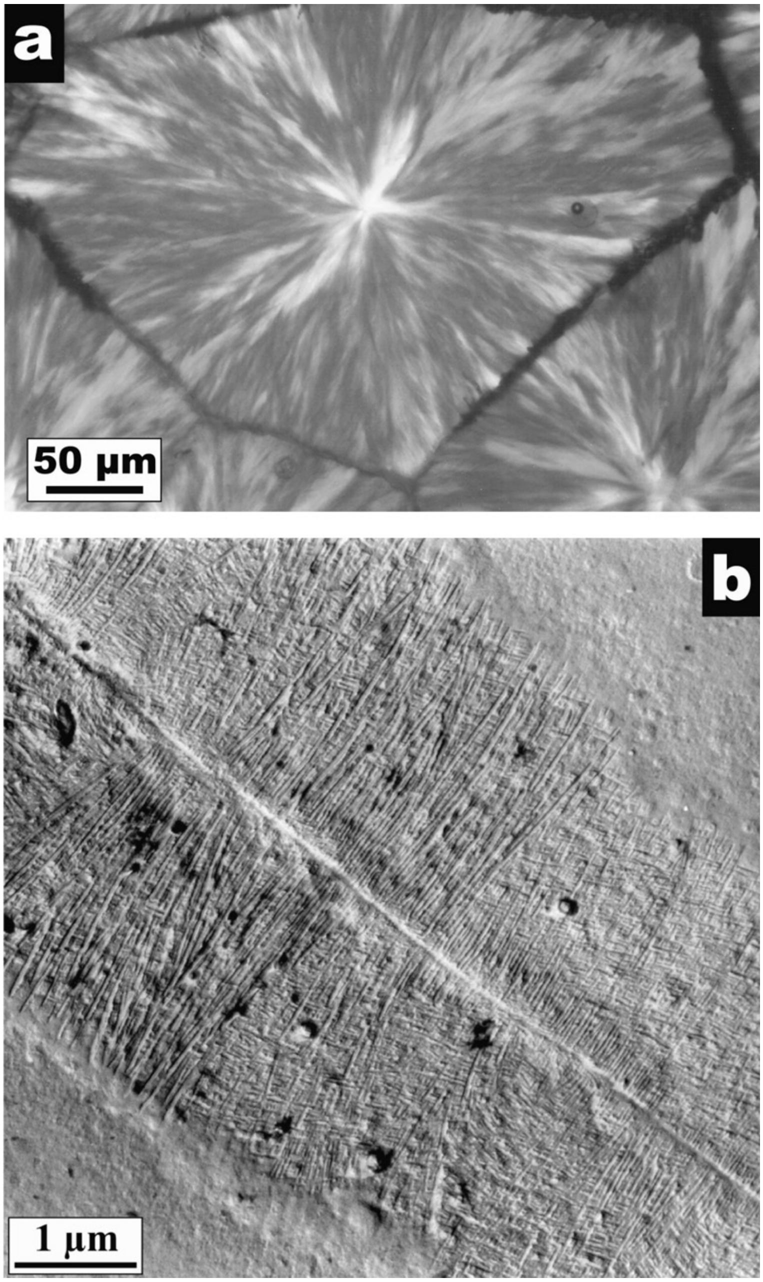


Figure 1
(a) A spherulite of i-polypropylene crystallized at 140 °C observed between crossed polars (and quarter-wave plate at 45°). This homopolymer had 93.5% pentad tacticity, a mass average molecular mass of 207,000 and a polydispersity of 1.5 (source: White (1995)). (b) A row structure of the same polymer crystallized at 140 °C with its lamellar structure revealed by permanganic etching (source: White (1995)).

as temperature falls. The radial texture varies with crystallization conditions and the particular polymer has all radii crystallographically equivalent (to the resolution of the polarizing microscope) when the "Maltese cross" extinction pattern between crossed polars remains stationary as the specimen is rotated in its own plane. Such extinction effects are well understood (Keith and Padden 1959, Price 1959, Keller

1959) but a stationary Maltese cross during rotation is ideal behavior from which most spherulites depart to a greater or lesser degree. It is common for regions which are not along the same radius to extinguish together, increasingly so for higher growth temperatures. However, this only serves to focus attention on the nature of the radial texture and to highlight the essential problem of spherulitic growth: how and why does a single crystal precursor develop into a spherical entity with more or less equivalent radii?

2. *Lamellar Microstructures*

A polymer spherulite (Fig. 2) is characteristically constructed on a framework of individual first-forming, so-called dominant lamellae, which advance separately and radially into the melt, repeatedly branching to fill space, with the new lamella diverging from its parent at a finite angle (commonly ~20° but increasing with supercooling, i.e., for thinner lamellae). Later-forming lamellae, termed subsidiary or infilling, grow from an existing dominant lamella and, typically, also diverge from it at similar angles. Crystallographic continuity is the norm but can be overridden when circumstances require. The balance between the advance of the open framework of dominant lamellae and its subsequent infilling is reflected in the kinetics of developing crystallinity as commonly analyzed according to the Avrami scheme (Phillips 1994). The finite divergence is not only responsible for the polycrystalline spread of growth directions in space but is evidently also the key phenomenon underlying spherulitic growth: repetition of divergence of dominant lamellae at successive branch points is sufficient to generate the morphology.

The origin of lamellar divergence is suggested by the early, sheaf-like forms seen particularly well at low supercoolings, in which lamellae with linear traces diverge regularly at approximately constant angles (Bassett 1984, Vaughan and Bassett 1989). These features are still present at higher supercoolings but in more branched and complex textures. Linear traces imply that lamellae grow without stress: they are very flexible and can be seen to deform readily when stressed. Whatever is responsible for the rather regular divergence must, therefore, only occur

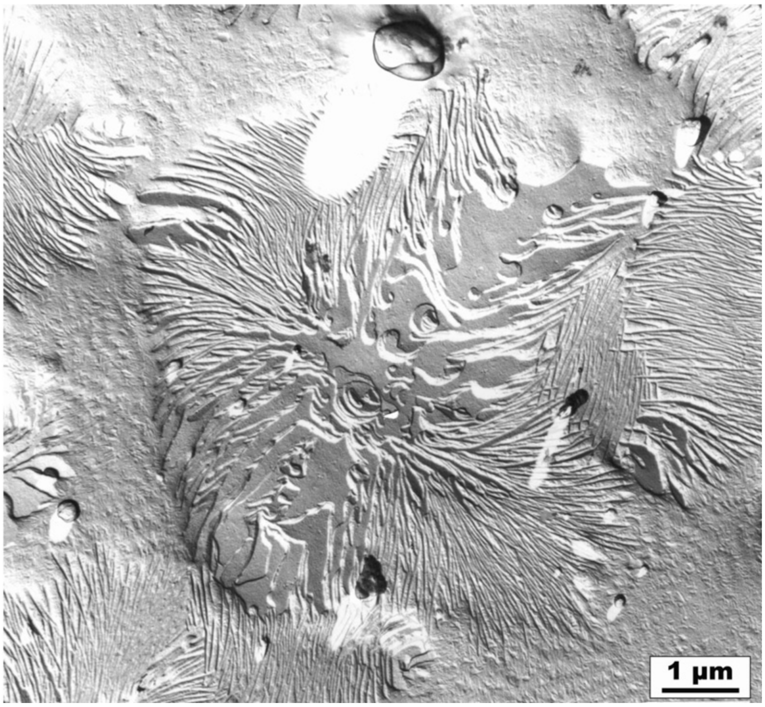


Figure 2
Lamellar construction of a linear-low-density polyethylene spherulite grown at 124 °C then quenched revealed by permanganic etching. Note the lamellar continuity from center to edge with new layers being introduced at (etched-out) giant screw dislocations. At the edge, individual dominant lamellae grow separately into the melt while later-developing subsidiary lamellae have appeared in the interior. Transmission electron micrograph of a shadowed carbon replica (source: Bassett (1994), Fig. 3).

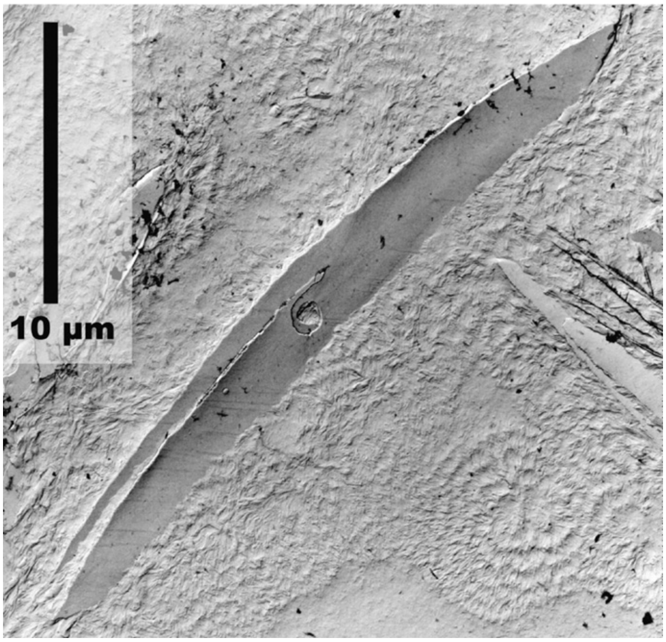


Figure 3
A single crystal of linear polyethylene growing around an (etched-out) giant screw dislocation within a quenched melt demonstrating the divergence of adjacent layers from their region of contact and the absence of stress away from the dislocation (unpublished).

close to the lamellar origin or branch point. The implication is that adjacent lamellae are being acted upon by a repulsive force but only in the neighborhood of their common contact, i.e., a force of mesoscopically short range.

Additional and unambiguous evidence of local strain, implying stress around the branch point, is seen in individual polyethylene crystals, within a quenched matrix, containing a giant screw dislocation (Vaughan and Bassett 1989, Bassett 1994). Layers of the associated spiral terrace, being continuous and part of the same crystal, instead of lying parallel in an unstrained state, are not in contact away from the dislocation core but strained so as to diverge with linear traces, making the respective growth directions, *b*, differently oriented in space (Fig. 3). The morphological evidence thus points firmly to the presence of a mesoscopic repulsion between adjacent lamellae as a primary cause of spherulitic growth in polymers.

The cause of the repulsion was first proposed to be pressure from uncrystallized portions of molecules partly attached to the crystal, i.e., cilia (Bassett 1984, 1994, Vaughan and Bassett 1989). This takes account of the distances involved being somewhat less than typical molecular lengths and extends the suggestion first put forward to account for rather similar splaying geometries in solution-grown crystals. Cilia will occupy space adjacent to lamellar surfaces, resisting any competition for that space with a weak rubbery modulus, probably higher for the shortest lengths. The recent availability of the monodisperse long *n*-alkanes has allowed a critical test of this longstanding hypothesis, leading to a very strong and detailed endorsement of the role of ciliation. It has also revealed the misaligned packing of initially rough fold surfaces to be a probable additional source of lamellar divergence.

3. Melt-crystallization of Monodisperse Long n-Alkanes

Early work on the first small quantities of such *n*-alkanes confirmed the expectation that they would tend to crystallize in quantized lamellar thicknesses with the methyl groups confined to the basal surfaces (Ungar *et al.* 1985). Two such thicknesses commonly occur from the melt, with fully-extended and once-folded molecular conformations. When crystallized from the melt as the once-folded form, these *n*-alkanes are spherulitic and may even exhibit banding, in which the optic axes spiral around the radius (Bassett *et al.* 1996a). On the one hand this is a clear demonstration that additives are not the cause of spherulitic growth; on the other hand the ciliation hypothesis predicts that, to the approximation that stems add to a lamella as a whole, only chainfolded and not extended-chain growth should give spherulites. This is because transient cilia will be present for the former when one fold stem has been laid down but never for the latter. The observed morphologies of $C_{294}H_{590}$, in the region of the changeover between once-folded and extended forms, match this prediction precisely (Bassett *et al.* 1996b). This is striking confirmation of the important role played by transient cilia in spherulitic growth, the details of which have been abundantly confirmed and illuminated by continuing studies of two types.

First, the meaning of transient cilia has been widened from studies of extended-chain growth in which, with increasing supercooling, dominant lamellae begin to branch and diverge with a substructure of parallel lamellae. To the initial concept of temporarily unattached portions of crystallizing molecules has been added the excess of molecular length over that of the secondary nucleus. This is in accord with plots of branching angle against supercooling, which are linear with positive slopes and intercepts. The latter implies, since transient ciliation must vanish at zero supercooling, that there must be an additional cause of lamellar divergence, suggested to be the misaligned packing of initially rough surfaces (Hosier and Bassett 2000).

Second, it is possible to introduce permanent cilia of controlled length and number in dilute cocrystallizing blends with a longer guest in a lamella of a shorter host (Hosier *et al.* 2000). These produce additional splaying, notably increasing the intercept—except when the guest molecule is twice the length of the host and able to crystallize once folded in the extended host without permanent cilia. A wealth of splaying data has been consistently related to permanent cilia plus misaligned packing of initially rough lamellar surfaces and transient ciliation, leaving no doubt as to the central roles of these phenomena in the spherulitic crystallization of long molecules.

In simple terms, such spherulites result because adjacent lamellae are either pushed apart near their contact or have initially rough surfaces which do not pack parallel. Equivalent mechanisms could readily be expected to apply to nonpolymeric systems such as ice, for which osmotic pressure has been suggested to result in splaying, although such systems are much less well characterized than polymers and alkanes.

4. *Spherulitic Textures*

Spherulites are not unique entities but have textures varying between one growth condition, one system and another. The subtlety of the variation has so far rendered all attempts at textural classification more or less inadequate, beginning with the coarse/fine and compact/open distinctions proposed for lateral textures and optical observations (Keith and Padden 1963). With knowledge of internal lamellar construction, the principal variables to quantify are identified as the angles and frequency of branching of dominant and subsidiary lamellae. Of these, only the angles by which dominant lamellae branch have been published: in poly(4-methylpentene-1) (Patel and Bassett 1994) and various *n*-alkane systems (Hosier and Bassett 2000, Hosier *et al.* 2000). These are among the simpler morphologies; for others, such as banded spherulites of polyethylene, textural quantification has yet to be achieved.

Isotactic polypropylene is a special case in that the unique cross-hatched morphology of the α or monoclinic form provides subsidiary lamellae at 80° to the dominants which change the optic sign of its spherulites from negative at high temperatures to positive at low temperatures because of the increasing frequency of cross-hatching in the latter. Data of cross-hatching frequency as functions of temperature and molecular weight are among those measured by White (1995).

The general change to a hedritic or axialitic morphology at higher crystallization temperatures is due, phenomenologically, to reduced branching. In polyethylene (Hoffman *et al.* 1975) this is more or less coincident with the change in dominant lamellar profiles from *S* to planar or ridged with $\{201\}$ surfaces (Bassett 1984, 1994, Vaughan and Bassett 1989). It has been suggested that, in this polymer, hedrites represent growth in kinetic regime I, when there is only one nucleus active on a growth surface, rather than many in the lower temperature regime II (Hoffman *et al.* 1975). More recently, the change to $\{201\}$ profiles has been shown to result from ordered fold packing becoming established before the next layer of fold stems is added to a lamella (Abo el Maaty and Bassett 2001a). Either or both eventualities can confidently be expected to reduce the frequency of branching, with the latter probably also reducing the magnitude, as smooth surfaces will pack better together.

5. *Segregation and Cellulation*

Spherulitic polymers are systematically spatially inhomogeneous with later-crystallizing and mobile noncrystallizing molecular species concentrated between dominants, and in the final regions to solidify. In polyethylene, shorter molecules tend to crystallize later in typical linear polymers but more branched ones in Ziegler–Natta copolymers (which correlate strongly with shorter chains). In polypropylene, slowly crystallizing short and less tactic molecules, together with atactic polymers, will be concentrated adjacent to the last sites of crystallization, including where three or more spherulites meet, possibly also with voids there if material becomes exhausted, with consequences for mechanical properties.

In exceptional conditions of high segregation of molecules which significantly lower the local isothermal supercooling, as in highly branched linear low-density polyethylenes, the IRG is no longer constant but falls monotonically towards a steady-state condition, promoting morphological instability and cellulation. Both occur after spherulitic growth is well established, once segregant concentrations reach sufficiently high values (Abo el Maaty *et al.* 1998). There is no recovery in growth rate back to the original value once the "impurity layer" has been breached, which was a (mistaken) premise of the original proposal (Keith and Padden 1963) that spherulitic textures derived from morphological instability. Segregation is an important consequence rather than the cause of spherulitic crystallization. It may also lead to cellulation independently of morphological instability, as noncrystallizing species locate periodically between advancing dominants. In a similar way, the lamellar advance of spherulites in segregating circumstances has been observed both optically and by atomic force microscopy not to occur smoothly: when lamellae in front are slowed, others overtake them from the rear, taking paths which avoid regions of high segregant concentration. Inherently different segregant concentrations produce differences in behavior between point and linear nucleation, spherulites, and row structures, respectively, including the faster growth of spherulites than rows in the same sample. In certain circumstances the IRG of row structures pulsates as a function of radial distance corresponding to repetitive changes in segregant concentration (Abo el Maaty and Bassett 2000).

6. Banded Growth

The spectacular phenomenon of banding in spherulites is not universal but restricted to particular systems, seemingly those in which lamellar normals are (ultimately) inclined to the molecular direction, i.e., systems in which the stress in {001} fold surfaces is high. The dark bands seen in the polarizing microscope are a consequence of zero birefringence due, as optical and x-ray examination showed long ago, to the averaged optical indicatrix spiralling around the radius; blackness occurs when an optic axis is parallel to the viewing direction. However, simple packing considerations show that it is impossible for all lamellae to spiral in this way.

Observation of the lamellar microstructure of polyethylene, the only major polymer showing banding systematically, shows that it has two kinds of spherulite (Hoffman *et al.* 1975, Abo el Maaty and Bassett 2001a). Above ~127 °C in the linear polymer they are unbanded and have dominant lamellae with {201} fold surfaces, a legacy of fold packing in a molecular layer having ordered before the next layer is added. Banded spherulites form at lower temperatures. Their dominant lamellae initially have perpendicular, i.e., {001}, fold surfaces, then develop *S*-profiles, viewed down *b*, with isothermal lamellar thickening and radial distance; the sense of the *S* is uniquely related to that of the overall twist. The complex morphology has dominant lamellae multiply connected with diagonal linkages based on giant screw dislocations, almost all isochiral, which occur systematically and preferentially to one side of an advancing lamellar tip. Between connections there is only moderate twisting; most of the overall twist is injected with an extra layer incrementally at the dislocations where the width of the ensuing lamella is also correspondingly much reduced (Vaughan and Bassett 1989).

Explanations for banding have, principally, to account for the source of the twisting stress and the consistent asymmetry of the system. One suggestion, based upon asymmetric deposition of molecular stems on an inclined lamella (Keith and Padden 1996), is contrary to unambiguous observation showing that when this occurs banding does not (Abo el Maaty and Bassett 2001a). When there is banded growth, molecules add to perpendicular lamellae, from which *S*-profiles develop as in the isolated crystal of Fig. 4. Here the *S* is most pronounced in the lamellar center and has its axis inclined to *b*, the growth direction (Bassett 1994, Patel and Bassett submitted). This asymmetry of habit is propagated by new lamellae inserted, at top and bottom, at isochiral giant screw dislocations.

In polyethylene, banding is driven by relief of fold surface stress and ordering of folds (Patel and Bassett submitted). The explanation offered is that, as fold packing orders in the center of the lamella, geometrical continuity will produce an *S*-profile and a torque ahead of it, eventually producing asymmetrically located, isochiral giant screw dislocations as the lamella yields, and associated diagonal linkages. Moreover, because the cross-section of the lamella is thereby reduced, the response to the torque will be correspondingly greater and give effectively an increment of twist to the new layer. This is a consistent explanation for both the complex morphological details of banding and its restriction to systems with inclined fold surfaces.

7. Properties of Spherulitic Morphologies

Diverging morphologies bring internal heterogeneity into spherulitic systems, on the scale of the interdominant spacing, typically ~1 μm or less, and thence an inherent spread of properties (Fig. 5). This heterogeneity may take many forms: geometrical, i.e., differences in local lamellar orientation, thickness, and crystallinity between dominant lamellae and subsidiary lamellae; with radial distance according to the extent of infilling growth; molecular fractionation by length, branch content, or tacticity, plus actual or incipient voids at interspherulite boundaries.

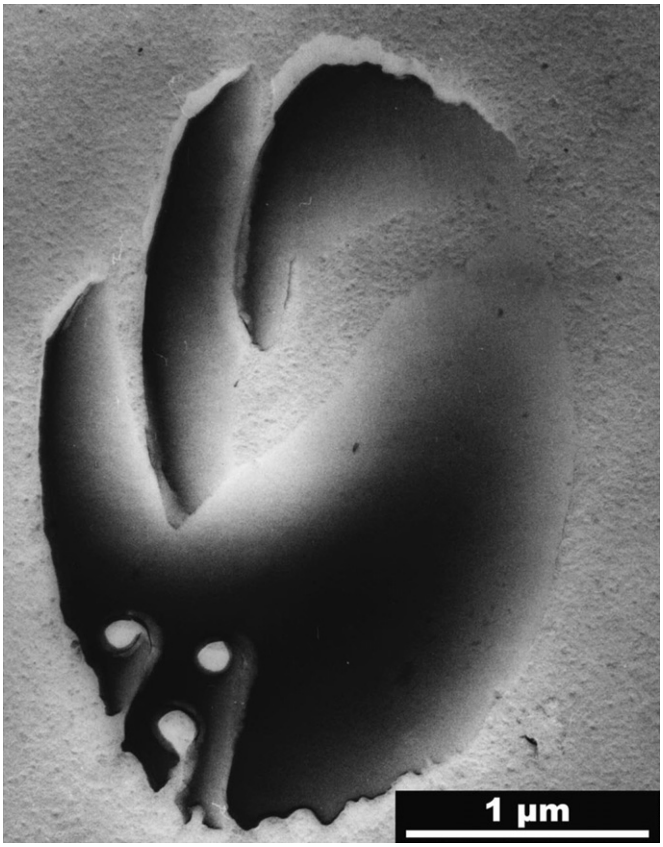


Figure 4
The development of an *S* profile in a polyethylene lamella. Note that axis of the *S* is inclined to the longest, *b*-axis, and the insertion of new layers to either side of the *S*-axis (source: Patel and Bassett (unpublished)).

The simplest situation has little or no fractionation as, for example, in a rapidly crystallized linear polyethylene, when the principal factors are finite spherulite size and the internal dominant/subsidiary morphology. The latter will cause refractive index differences, at least on the scale of interdominant separations, together with wider variations for nonparallel dominants. These are likely to cause radial streaking in the optical microscope, especially when phase-contrast effects operate in slightly out-of-focus images. More importantly, they and whole spherulites will contribute to Rayleigh scattering and so adversely affect optical clarity. It follows that a high nucleation density is desirable to reduce scatter in crystalline polymers. In systems such as i-polypropylene, where segregation is more complex, greater differences are to be expected.

In terms of mechanical properties, the orientation of lamellae with respect to the applied stress is a most important influence on their plastic deformation (Bassett 2000). As this differs substantially between a dominant lamella and its neighboring subsidiaries, they are very liable to deform by different mechanisms (Fig. 6). For example, in tensile drawing, there will be no shear stress parallel to the *c*-axis in dominant lamellae parallel to the draw direction but this will not be the case for their subsidiaries inclined at 20–30°. The latter should, therefore, be prone to yield in shear before the adjacent dominants, in line with observations of the sites of localized deformation. Additional reasons for differential deformation may arise when subsidiary lamellae are thinner than dominants.

Until very recently received opinion was that notwithstanding the manifold complications of spherulitic deformation this history was lost in tensile cold drawing. This is not the case. Recent work has shown that, in linear polyethylene, transverse memory remains even in samples drawn as much as ~ 50 times (Amornsakchai *et al.* 2000, 2001). This novel finding complements much previous work which concentrated almost exclusively on longitudinal morphology, often by diffraction methods. Related observations are the systematic transverse morphologies in "overdrawn" polypropylene and advanced fibers of

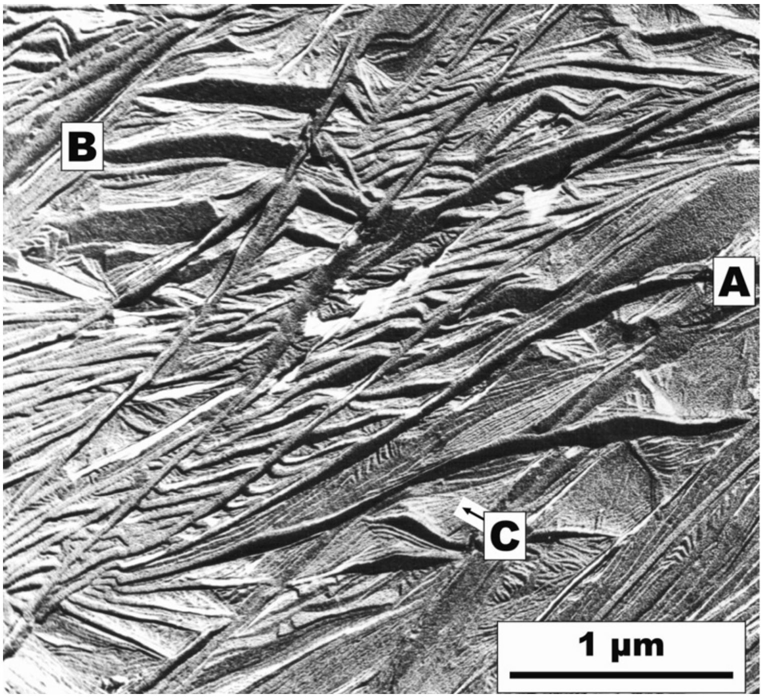


Figure 5
The interior of a linear polyethylene spherulite crystallized at 128 °C demonstrating the long and diverging dominant lamellae (A), the shorter subsidiary lamellae (B) growing from them, and infilling lamellae (C), which in this instance formed on quenching (source: Freedman *et al.* (1986) *Polymer* **27**, 1163–9, Fig. 2).

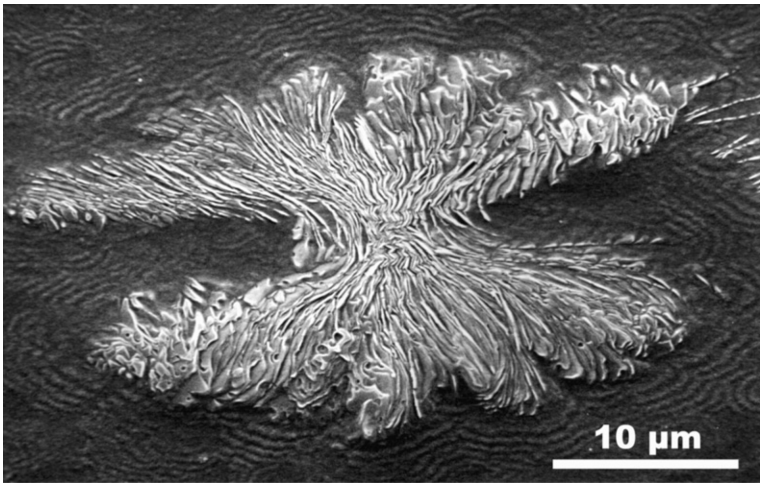


Figure 6
An immature spherulite of linear polyethylene (in a linear-low-density matrix) drawn to 50% extension at 100 °C reveals the different modes of deformation according to the angle of a lamella to the horizontal tensile axis. Those nearly parallel to the axis have broken up while those perpendicular to it have rotated cooperatively to a herringbone conformation (source: Lee *et al.* (2000) *J. Mat. Sci.* **35**, 5101–10, Fig. 3b).

polyethylene, leading to the inference of a new model for fiber structure (Bassett 2000).

The progressive manner in which lamellae crystallize in a spherulite is liable to leave a legacy of differential melting points in the morphology. In conventional linear low-density polyethylene, the melting range is broad because of the successive crystallization of increasingly branched molecules whose excluded branches limit the lamellar thickness to thinner and thinner dimensions with correspondingly lowered melting points. The locations of low melting material can be located by annealing the material into the melting range—which liquefies that part of the sample melting below the annealing temperature—then identifying changes from the original morphology within the annealed sample. Study of poly(4-methylpentene-1) in this way has not only demonstrated the lower melting points of subsidiary lamellae but has also identified the occurrence of isothermal lamellar thickening in lamellae recrystallized after initial melting (Bassett 1994).

Differential melting is a major feature of the unique cross-hatched morphology of monoclinic i-polypropylene. For typical isothermal crystallization temperatures and molecular lengths, cross-hatching lamellae are thinner than the dominant radial lamellae in spherulites with a correspondingly lower melting point. At annealing temperatures between the respective melting points, all cross-hatching lamellae melt and, on recrystallization, form new radial lamellae. The frequency of cross-hatching increases for lower crystallization temperatures and molecular lengths (White 1995) which has the effect of altering the overall sign of the initial birefringence when the balance between the opposing contributions of radial and cross-hatched lamellae reverses. There is also a newly discovered effect in the absence of cross-hatching. Dominant lamellae crystallized at 160 °C, when cross-hatching does not occur, melt ~3 K higher than intervening lamellae (Weng *et al.* in press) but only when they are separated by more than molecular lengths. In spherulite centers this is not the case due, it is suggested, to internuclear interference (Abo el Maaty and Bassett 2001b) and correspondingly the melting point is reduced. This difference in properties between central and outer regions is most important because the balance between the two is controlled by nucleation density: commercially important differences arise in this way.

8. Spherulites and Processing

Spherulitic growth is characteristic of quiescent systems and is appropriately modified in more commercial processes involving flow crystallization with temperature gradients and/or applied stress. Tensile stress applied to a molten polymer may be sustained for short times by transient entanglements which transmit the stress through the sample. In so doing they produce elongational strain in highly aligned regions capable of acting as linear nuclei, leading to the formation of row structures. When linear nuclei are present, isothermal crystallization occurs at higher supercoolings because of the rise in melting point with extension of the melt, and lamellar growth is affected by the applied stress. The interpretation of such morphologies to indicate local crystallization conditions benefits from comparison with stress-free isothermal crystallization of rows using fibers as nuclei. This has shown not only that lamellar habit changes with radial distance (Abo el Maaty and Bassett 2001a) (hence nucleation density and applied stress) but also new phenomena attributed to internuclear interference (Abo el Maaty and Bassett 2001b) in which different nuclei on adjacent lamellae compete for the same molecule, thereby slowing both the IRG and isothermal lamellar thickening. Internuclear interference is a new concept which, *inter alia*, can account for the existence of dominant lamellae even in extended-chain lamellae of the long *n*-alkanes and thence the inherent spatial inhomogeneity of crystallized long molecular systems.

9. Conclusions

The two major conclusions underlying polymer spherulites and their associated properties are, first, that spherulites form because dominant lamellae branch then diverge repeatedly. For long molecules the divergence results from ciliation and/or misalignment of initially rough surfaces. Second, the presence of diverging lamellae gives an inherent spatial modulation of properties, which can be bimodal, because dominant lamellae tend to behave differently from intervening ones crystallized later, in more restricted environments, which are liable to be both differently oriented and thinner.

Bibliography

Abo el Maaty M I, Bassett D C 2000 On pulsating growth rates in banded crystallization of polyethylene. *Polymer*. **41**, 9169

Abo el Maaty M I, Bassett D C 2001a On fold surface ordering and reordering during the crystallization of polyethylene from the melt. *Polymer* **42**, 4957

Abo el Maaty M I, Bassett D C 2001b On interfering nuclei and their novel kinetics during the crystallization of polyethylene from the melt. *Polymer* **42**, 4965

Abo el Maaty M I, Hosier I L, Bassett D C 1998 A unified context for spherulitic growth in polymers. *Macromolecules* **31**, 153

Amornsakchai T, Olley R H, Bassett D C, Al-Hussein M O M, Unwin A P, Ward I M 2000 On the influence of initial morphology on the internal structure of highly-drawn polyethylene. *Polymer* **41**, 8291

Amornsakchai T, Bassett D C, Olley R H, Unwin A P, Ward I M 2001 Remnant morphologies in highly-drawn polyethylene after annealing. *Polymer* **42**, 4117

Bassett D C 1984 Electron microscopy and spherulitic organization in polymers. *CRC Crit. Rev. in Solid State and Materials Sciences* **12**, 97

Bassett D C 1994 Lamellae and their organization in melt-crystallized polymers. *Phil. Trans. R. Soc. London, A* **348**, 29

Bassett D C 2000 Deformation mechanisms and morphology of crystalline polymers. In: Ward I M, Coates P D, Dumoulin M M (eds.) *Solid Phase Processing of Polymers. Hanser, Munich*, p. 11

Bassett D C, Olley R H, Sutton S J, Vaughan A S 1996a On spherulitic growth in a monodisperse paraffin. *Macromolecules* **29**, 1852

Bassett D C, Olley R H, Sutton S J, Vaughan A S 1996b On chain conformations and spherulitic growth in monodisperse n-$C_{294}H_{590}$. *Polymer* **37**, 4993

Hoffman J D, Frolen L J, Ross G S, Lauritzen J I 1975 On the growth rate of spherulites and axialites from the melt in polyethylene fractions: regime I and regime II crystallization. *J. Res. Natl. Bur. Stds. A* **79**, 671

Hosier I L, Bassett D C 2000 A study of the morphologies and growth kinetics of three monodisperse n-alkanes: $C_{122}H_{246}$, $C_{162}H_{326}$ and $C_{246}H_{494}$. *Polymer* **41**, 8801

Hosier I L, Bassett D C, Vaughan A S 2000 Spherulitic growth and cellulation in dilute blends of monodisperse long n-alkanes and their blends. *Macromolecules* **33**, 8781

Keith H D, Padden F J 1959 The optical behavior of spherulites in crystalline polymers. I: Calculation of theoretical extinction patterns in spherulites with twisting crystalline orientation. II: The growth and structure of the spherulites. *J. Polym. Sci.* **39**, 101, 123

Keith H D, Padden F J 1963 A phenomenological theory of spherulitic crystallization. *J. Appl. Phys.* **34**, 2409

Keith H D, Padden F J 1996 Banding in polyethylene and other spherulites. *Macromolecules* **29**, 7776

Keller A 1958 Morphology of crystalline polymers. In: Doremus R H, Roberts B W, Turnbull D (eds.) *Growth and Perfection of Crystals*. Wiley Interscience, Chichester, UK, p. 499

Keller A 1959 Investigations on banded spherulites. *J. Polym. Sci.* **39**, 151

Patel D, Bassett D C 1994 On spherulitic crystallization and the morphology of melt-crystallized isotactic poly(4-methylpentene-1). *Proc. R. Soc. London, A* **445**, 577

Patel D, Bassett D C submitted On the formation of S-profiled lamellae in polyethylene and the genesis of banded spherulites.

Phillips P J 1994 Spherulitic crystallization in macromolecules. In: Hurle D T J (ed.) *Handbook of Crystal Growth*. North-Holland, New York, Vol. 2, p. 1167

Price F P 1959 On extinction patterns of polymer spherulites. *J. Polym. Sci.* **39**, 139

Ungar G, Stejny J, Keller A, Bidd I, Whiting M C 1985 The crystallization of ultralong normal paraffins: the onset of chainfolding. *Science* **229**, 386

Vaughan A S, Bassett D C 1989 Crystallization and morphology. In: Booth C, Price C (eds.) *Comprehensive Polymer Science-2*. Pergamon, Oxford, UK, p. 415

Weng J, Bassett D C, Olley R H, Jääskeläinen P, in press On morphology and multiple melting in polypropylene. *J. Macromol. Sci. Phys.*

White H M 1995 Lamellae and their organization in melt-crystallized isotactic polypropylene. Ph.D. thesis, University of Reading, UK

D. C. Bassett
University of Reading, UK

Polymer Surfaces: Structure

Since thin films of polymeric materials are used in a broad variety of technological applications, e.g., paints, lubricants, and adhesives, a critical characterization of their thermophysical properties is essential. Examples of the importance of such characterization include work on chain diffusion (Zheng *et al.* 1997) and glass transition phenomena (Keddie *et al.* 1994, van Zanten *et al.* 1996, Forrest *et al.* 1996) in thin polymer films. In both cases it has been suggested that the modification of chain conformations on confinement may play an important role. Similarly, phase transitions in thin polymer mixtures (Kumar *et al.* 1994, Rouault *et al.* 1995) can also be affected by the alteration of chain conformation. Brown and Russell (1996) have suggested that modifications of chain conformation in the vicinity of surfaces can alter the interchain entanglement density in this region and thus explain the modification of system properties.

There has been considerable effort since the mid-1980s towards characterizing the structure and conformations of thin films of amorphous polymers using theory, experiment, and simulations. Developments in this area are outlined first. The consequence of these findings on system properties, especially their dynamics, is also discussed. There has also been some experimental work in the area of liquid crystalline polymers, such as those relevant to liquid crystal displays, where these polymers play the role of an orienting layer. These results are discussed, but almost no mention of theory is made, since this aspect has received little attention in the published literature. Finally, emerging results on thin films of crystalline polymers are addressed.

1. *Amorphous Polymers*

Conjectures on the structure and conformation of thin films of amorphous polymers can be traced to the work of deGennes (1979) and Silverberg (1982). In these and many other theoretical works it has been assumed that chain conformations parallel and perpendicular to the surfaces behave very differently. In the confinement direction (perpendicular to the surfaces) chains assume smaller dimensions, as expected. However, in the direction parallel to the surfaces it has been assumed that the chains maintain their unperturbed Gaussian conformations. While these are conjectures based on the simple fact that a polymer melt chain assumes a Gaussian conformation with three decoupled directions, they have been very popular since properties change only along the confinement direction, thereby affording considerable simplification. These assertions have, somewhat surprisingly, been validated by computer simulations on lattice (Pakula 1991, Yethiraj and Hall 1991, Mansfield and Theodorou 1991, Bitsanis and ten Brinke

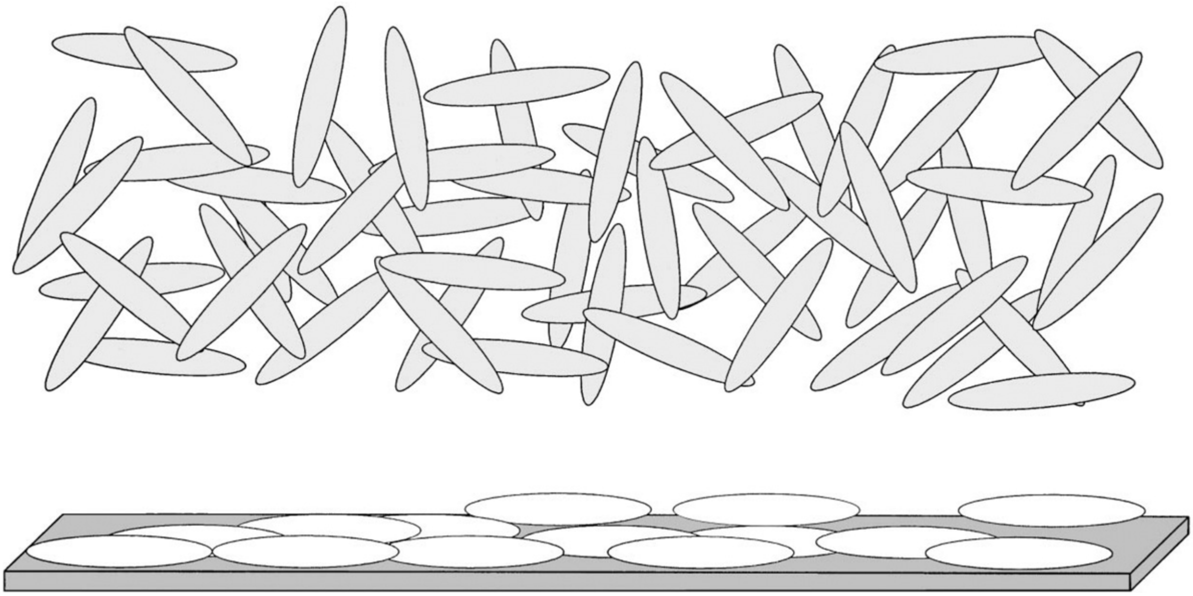


Figure 1
Schematic of instantaneous chain conformation in the bulk and near surfaces for amorphous polymers.

1993) and off-lattice models (Kumar *et al.* 1988) with various degrees of coarse graining. The internally consistent picture that emerges from these simulations is shown schematically in Fig. 1. The instantaneous conformation of a polymer chain in the bulk can be represented as an ellipsoid. Since these ellipsoids are free to rotate in space, they appear on average to yield the well-known spherical shape for a melt chain. When a melt is confined into a thin film, especially when the film thickness is smaller than the longest axis of the ellipsoid, the chains will reorient so as to place their shortest axis parallel to the confinement direction. This is preferred since a chain will then only lose $k_B T$ of free energy, corresponding to the loss of one rotational degree of freedom. Thus, the chain dimensions in the direction parallel to the surfaces represents an average of the two longest axes of the ellipsoid, which turns out to be different from the melt value by less than 10%. Further, the chain conformation will remain essentially Gaussian in this direction. Of course, this picture will change rather dramatically when chains are confined into films thinner than the shortest axis of the ellipsoid. In this situation one expects that chains will be perturbed strongly, and possibly no longer assume their Gaussian conformations.

The conformation of polymer chains in films over a broad range of thickness ($0.5 \leqslant D/R_g \leqslant \infty$, where D is the film thickness and R_g is the unperturbed radius of gyration) have been measured using small-angle x-ray and neutron scattering (Shuto *et al.* 1993, Frank *et al.* 1996) and grazing incidence neutron scattering. Since the scattering vectors in these experiments are predominantly in the plane of the film, they are primarily sensitive to chain dimensions in this direction alone. In general, films of isotopically labeled polymers are created by spin coating or by Langmuir–Blodgett deposition. These *unannealed* samples are then examined by scattering. It is found that chain dimensions in the plane of the film increase with decreasing film thickness, especially in the range $D/R_g \leqslant 1$. However, since these samples are not annealed, they are unlikely to be at equilibrium. Hence, these data are probably not the most suitable to examine critically simulation results and theoretical conjectures on the conformation of chain molecules in confined geometries. Studies (Jones *et al.* 1999) on annealed polymer films have found that chains assume their unperturbed Gaussian conformations in the direction parallel to the surfaces, to within an experimental uncertainty of 10%, independent of film thickness, i.e., for $D/R_g \geqslant 0.5$. These results are consistent with the predictions of simulation and theory. In this context it must be emphasized that the experiments reported above correspond to a single polymer, polystyrene. More experiments are needed on other polymers, and on different substrates, to verify the generality of these results.

The consequences of these ideas for the properties of thin polymer films are now examined. It appears to be well established that the density of films as thin as $D/R_g = 0.5$ are essentially the same as the bulk value (Wallace *et al.* 1998). Brown and Russell (1996) have suggested that the constraints of bulk-like density

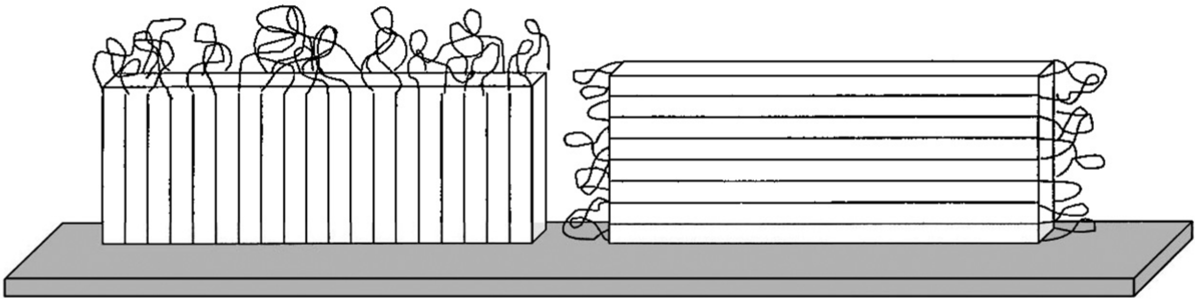


Figure 2
Schematic of two possible models of polymer crystallization in ultrathin films.

and an essentially Gaussian conformation with chains oriented with their long axes parallel to the surface imply that these chains have lowered intermolecular entanglement. This implies that confinement will have dramatic consequences on the dynamics and mechanical properties of thin films relative to their bulk counterparts. Such conclusions, indeed, appear to be in qualitative agreement with the diffusion and glass transition phenomena in thin films. However, this area is still in its nascent stage and the connection between chain structure and properties is not yet fully understood.

2. *Liquid Crystalline Polymers*

The near-surface structure of liquid crystalline (LC) polymers is important in thin-film display technology. In these situations a thin film of a LC polymer is coated on a substrate. This layer is then uniaxially "buffed." Small-molecule nematogens placed on this buffed surface orient in the buffing direction. Since the orientation of the nematogens is driven by the local orientation of the buffed LC polymer chains, much effort has been directed toward characterizing the near-surface structure of these films. Techniques employed include grazing incidence x-ray scattering, which provides information on in-plane structure as a function of depth, and near-edge x-ray absorption fine structure (NEXAFS) (Samant *et al.* 1997), which provides information on the orientation of molecular groups either in the top 1 nm or the top 10 nm, depending on the detection technique employed. These techniques have been complemented by bulk characterization techniques such as IR spectroscopy (Heitpas and Allara 1998).

Through these approaches it has been established that the near surface of poly(pyromelliticdianhydride-oxydianiline) (PMDA-ODA), which is formed by *in situ* imidization of the amic acid precursor, has enhanced ordering as compared to the bulk, over distances of 10 nm from the surface. This is a result for a single polymer, and hence its applicability to a broad class of LC polymers remains an open question.

In the case of poly(biphenyldianhydride-*p*-phenylenediamine) (BPDA-PDA), it has been found that the chain segments in the top 1–10 nm are preferentially oriented in the buffing direction. The molecular origins of this orientation have been tentatively identified as being owing to a surface plastic deformation, driven by the high local stress during the buffing. Such results have been observed in a wide class of polymers, including common polymers such as polystyrene, thus illustrating that polymer surfaces can be oriented by what appears to be only mild rubbing conditions.

3. *Crystalline Polymers*

The behavior of crystalline polymers near surfaces and in thin films has received little study. Work (Frank *et al.* 1996, Prucker *et al.* 1998) appears to suggest that these materials can crystallize as long as films are much thicker than 40 nm. However, below this thickness the material will no longer crystallize. A possible scenario for this phenomenon is sketched in Fig. 2, where it is shown that if chain stems orient normal to the surface a minimum film thickness for which crystallization can occur corresponds roughly to the long period of these systems. These are purely conjectures at this time since it is unclear if chain stems do orient in this direction, and if the lamellar periods remain effectively unchanged even in these thin films.

Bibliography

Auroy P, Auvray L, Leger L 1991 Characterization of the brush regime for grafted polymer layers at the solid–liquid interface. *Phys. Rev. Lett.* **66**, 719–22

Bitsanis I, ten Brinke G 1993 A lattice Monte Carlo study of long chain conformations at solid–polymer interfaces. *J. Chem. Phys.* **99**, 3100–11

Brown H R, Russell T P 1996 Entanglements at polymer surfaces and interfaces. *Macromolecules* **29**, 798–800

deGennes P G 1979 *Scaling Concepts in Polymer Physics*. Cornell University Press, Ithaca, NY

DeMaggio G B, Frieze W E, Gidley D W, Zhu M, Hristov H A, Yee A F 1997 Interface and surface effects on the glass transition in thin polystyrene films. *Phys. Rev. Lett.* **78**, 1524–7

Forrest J A, Dalnoki-Veress K, Dutcher J R 1997 Interface and chain confinement effects on the glass transition temperature of thin polymer films. *Phys. Rev. E* **56**, 5705

Forrest J A, Dalnoki-Veress K, Stevens J R, Dutcher J R 1996 Effect of free surfaces on the glass transition temperature of thin polymer films. *Phys. Rev. Lett.* **77**, 2002–5

Frank C W, Rao V, Despotopoulou M M, Pease R F W, Hinsberg W D, Miller R D, Rabolt J F 1996 Structure in thin and ultrathin spin-cast polymer films. *Science* **273**, 912–5

Heitpas G, Allara D 1998 Molecular structure of poly(biphenyl dianhydride-*p*-phenylenediamine) polyamide thin films by infrared spectroscopy: thickness dependence of the structure in the nano- to micrometer range. *J. Polym. Sci. A: Polym. Chem.* **36**, 1247

Heitpas G, Sands J, Allara D 1998 A vibrational spectroscopic study of molecular restructuring at surfaces of unidirectionally rubbed polyimide thin films. *J. Phys. Chem. B* **102**, 10556

Jones R L, Kumar S K, Ho D L, Briber R M, Russell T P 1999 Chain conformation in ultrathin polymer films. *Nature* **400**, 146–9

Keddie J L, Jones R A L, Cory R A 1994 Size-dependent depression of the glass transition temperature in polymer films. *Europhys. Lett.* **27**, 59–64

Kerle T, Yerushalmi-Rozen R, Klein J 1997 Cross-link-induced autophobicity in polymer melts: a re-entrant wetting transition. *Europhys. Lett.* **38**, 207–12

Kikuchi H, Sands J, Kumar S K 1995 Near-surface alignment in rubbed polymer films. *Nature* **374**, 709–11

Kosmas M K 1990 Ideal polymer chains of various architectures at a surface. *Macromolecules* **23**, 2061–5

Kumar S K, Tang H, Szleifer I 1994 Phase transitions in thin films of symmetric polymer mixtures. *Mol. Phys.* **81**, 867–72

Kumar S K, Vacatello M, Yoon D Y 1988 Off-lattice Monte Carlo simulations of polymer melts confined between two plates. *J. Chem. Phys.* **89**, 5206–15

Mansfield K, Theodorou D N 1991 Molecular dynamics simulation of a glassy polymer surface. *Macromolecules* **24**, 6283–94

Mansky P, Liu Y, Huang E, Russell T P, Hawker C 1997 Controlling polymer–surface interactions with random copolymer brushes. *Science* **275**, 1458–60

Mayes A M, Russell T P, Bassereau P 1994 Evolution of order in thin block copolymer films. *Macromolecules* **27**, 749–55

Pakula T 1991 Computer simulation of polymers in thin layers: I. Polymer melt between neutral walls—static properties. *J. Chem. Phys.* **95**, 4685–90

Prucker O, Christian S, Bock H, Ruhe J, Frank C W, Knoll W 1998 On the glass transition in ultrathin polymer films of different molecular architecture. *Macromol. Chem. Phys.* **199**, 1435–44

Rouault Y, Baschnagel J, Binder K 1995 Phase separation of symmetrical polymer mixtures in thin film geometry. *J. Stat. Phys.* **80**, 1009–31

Samant M, Stohr J, Brown H R, Russell T P, Sands J M, Kumar S K 1997 NEXAFS studies on the surface orientation of buffed polyimides. *Macromolecules* **29**, 8334–42

Shuto K, Oishi Y, Kajiyama T 1995 Non-equilibrium characteristics of a two-dimensional ultrathin film prepared by the water casting method. *Polymer*. **6**, 549–57

Shuto K, Oishi Y, Kajiyama T, Han C C 1993 Aggregation structure of a two-dimensional ultrathin polystyrene film prepared by the water casting method. *Macromolecules* **26**, 6589–94

Sikka M, Cerini L N, Ghosh S S, Winey K I 1996 Melt intercalation of polystyrene in layered silicates. *J. Polym. Sci. Polym. Phys. Ed.* **34**, 1443–9

Silverberg A 1982 Distribution of conformations and chain ends near the surface of a melt of linear flexible macromolecules. *J. Coll. Interface Sci.* **90**, 86

Theodorou D N 1988 Microscopic structure and thermodynamic properties of bulk copolymers and surface-active polymers at interfaces: 1. Theory. *Macromolecules* **21**, 1411–21

Vaia R A, Giannelis E P 1997 Polymer melt intercalation in organically modified layered silicates: model predictions and experiment. *Macromolecules* **30**, 8000–9

van Zanten J H, Wallace W E, Wu W L 1996 Effect of strongly favorable substrate interactions on the thermal properties of ultrathin polymer films. *Phys. Rev. E* **53**, R2053–6

Wallace W E, Tan N C B, Wu W L, Satija S 1998 Mass density of polystyrene thin films measured by twin neutron reflectivity. *J. Chem. Phys.* **108**, 3798–804

Yethiraj A, Hall C K 1991 Square-well chains: bulk equation of state using perturbation theory and Monte Carlo simulations of the bulk pressure and of the density profile near walls. *J. Chem. Phys.* **95**, 1999–2005

Zhao X, Zhao W, Zheng X, Rafailovich M H, Sokolov J, Schwarz S A, Pudensi M A A, Russell T P, Kumar S K, Fetters L J 1992 Configuration of grafted polystyrene chains in the melt: temperature and concentration dependence. *Phys. Rev. Lett.* **69**, 776–9

Zheng X, Rafailovich M, Sokolov J, Strzhemechny Y, Schwarz S A, Sauer B B, Rubenstein M 1997 Long-range effects on polymer diffusion induced by a bounding interface. *Phys. Rev. Lett.* **79**, 241–4

S. K. Kumar and R. L. Jones
Pennsylvania State University, University Park,
Pennsylvania, USA

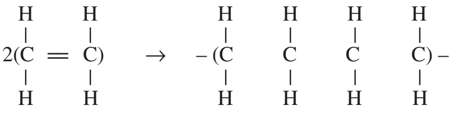
Polymer–Ceramic Nanocomposites: Polymer Overview

Macroscopic inorganic–organic hybrids, notably fiberglass-reinforced epoxy, are well accepted composite materials. While they are essential to many industries, such as electronics, aerospace, and automotive, macroscopic composites have lost some of their glamour. Now the excitement in advanced materials is nanocomposites. Nanocomposites refer to the family of materials where the inorganic and organic components are mixed on the nanometer scale. This first section introduces the polymers that have been mixed in various ways with inorganic materials, mainly zeolites, clays, ceramics, and glasses. The details of synthesis and the discussions of structure–property relationships are left to the sections that follow.

This introduction includes both (i) simple textbook definitions of terms used in polymer technology and (ii) a “roadmap” to the literature on nanocomposites. Principal references and source materials in this field include special issues of the *Journal of Sol–Gel Science and Technology* (1995, 1996), Drzal *et al.*

(1995), Coltrain *et al.* (1996), Laine *et al.* (1998), Klein *et al.* (1999) and Andrews and Najafi (1997).

By way of definition, a polymer is a long-chain molecule, a macromolecule, usually with a carbon backbone. The building block is the monomer. The polymer increases in molecular weight either by (i) addition polymerization or (ii) condensation polymerization. Addition polymerization involves the reaction of breaking a double bond to form two single bonds. A typical addition polymerization for ethylene is:



$$2\begin{pmatrix}\text{H} & & \text{H}\\ | & & |\\ \text{C} & = & \text{C}\\ | & & |\\ \text{H} & & \text{H}\end{pmatrix} \rightarrow -\begin{pmatrix}\text{H} & \text{H} & \text{H} & \text{H}\\ | & | & | & |\\ \text{C} & \text{C} & \text{C} & \text{C}\\ | & | & | & |\\ \text{H} & \text{H} & \text{H} & \text{H}\end{pmatrix}- \tag{1}$$

with the application of heat or light. Typical addition polymers are polyethylene, poly(vinyl chloride) (PVC), poly(vinyl acetate) (PVAc), poly(methyl methacrylate) (PMMA), and polystyrene.

Condensation polymerization involves the reaction of two small molecules and produces a nonpolymerizable by-product. A polymerization for nylon is:

$$\underset{\text{Aminoacid}}{2\text{NHRR}'\text{COOH}} \rightarrow \underset{\text{“B”}}{\text{NHRR}'\text{CO}-\text{NRR}'\text{COOH}} + \text{H}_2\text{O} \tag{2a}$$

$$\text{“B”} + \text{NHRR}'\text{COOH} \rightarrow$$
$$\text{NHRR}'\text{CO}-\text{NRR}'\text{CO}-\text{NRR}'\text{COOH} + \text{H}_2\text{O} \tag{2b}$$

where R is a univalent radical and R' is a divalent radical ($(CH_2)_n$ where $n \geqslant 5$). The first step is carried out at 175 °C with a catalyst and the second step is carried out at 250 °C under vacuum. Typical condensation polymers include nylon (hexamethylamine and adipic acid), polyamides, polyimides, and polyesters.

The functionality of the polymer refers to the number of reaction sites available. A bifunctional monomer produces linear polymers. A trifunctional or tetrafunctional monomer produces network polymers. When polymerization occurs between two different monomers, the result is a copolymer.

A reasonable question to ask at this point is why combine organic and inorganic materials? Some obvious reasons are that (i) organic materials in inorganic ones reduce the overall density of the object, (ii) inorganic materials in organic ones improve their hardness and mechanical properties in elastomeric, glassy, and semicrystalline states, (iii) organic polymers in porous inorganics produce pore-free materials without high-temperature treatment, and (iv) mixing inorganic and organic materials on the nanoscale introduces new phenomena we are just beginning to understand.

In the following sections materials are described that are designed, in some cases, for structural applications and, in other cases, for functional uses, for optical, electronic, and catalytic applications. The processes used to prepare these materials variously lead to one-, two-, and three-dimensional structures. The phases are described as "guest and host," "infiltrant and matrix," and *in situ*. The common theme is that all of the materials are classified as hybrids and all of them are nanocomposites.

1. *Definition of Nanocomposite*

First, we use the term hybrid as the generic term for an inorganic–organic material. An accepted name for many of these materials is organically modified silicates (ormosils) (Schmidt and Wolter 1990) or, more generally, organically modified ceramics (ormocers) (Schmidt 1994). Another name is ceramic modified with polymer (ceramer) (Huang *et al.* 1985).

The second term we use commonly is nanocomposite to emphasize the nanometer level of mixing. Nanocomposites often refer to hybrids where organic and inorganic constituents are not covalently bonded, but it also applies to covalently bonded constituents. The forces acting between the constituents vary from weak to relatively strong.

Some notable attempts to classify hybrids have appeared. For example, Chujo and Saegusa (1992) have reviewed inorganic–organic materials prepared where the organic constituents are polymers, but not low molecular weight radicals or groupings. Sanchez and Ribot (1994) divide hybrids into two classes: Class I corresponds to hybrids where organic molecules are blended into the inorganic network, whereas Class II includes hybrids where inorganic and organic constituents are linked together via covalent bonds. Avnir *et al.* (1998) concentrate on inorganic–organic hybrids using organofunctional silicon alkoxides. Wen and Wilkes (1996) classify the resulting materials as networks. The dimensionality of hybrids has been reviewed by Lerner and Oriakhi (1997), principally in layered materials. We prefer the terms physical hybrids and chemical hybrids (Wojcik and Klein 1997). Overall these classifications organize hybrids into groups by looking at (i) the structure, (ii) methods of preparation, or (iii) the nature of links between organic and inorganic constituents.

The majority of processes treated in the following sections refer to the sol–gel process (Brinker and Scherer 1990). The sol–gel process is defined in this context as the hydrolytic process with a metal alkoxide such as tetraethylorthosilicate (TEOS). Upon addition of water TEOS in a cosolvent hydrolyzes:

$$Si(OC_2H_5)_4 + 4H_2O \rightarrow Si(OH)_4 + 4C_2H_5OH \tag{3}$$

and polymerizes:

$$Si(OH)_4 + Si(OH)_4 \rightarrow (OH)_3Si-O-Si(OH)_3 + H_2O \tag{4}$$

to build a silicate network. In time, the system arrives at the sol–gel transition where an irreversible gel exists. In the single component system, it is known that the result is an oxide skeleton which is porous once the solvent is removed.

The sol–gel approach as described is a room-temperature process. Polymers that otherwise are destroyed at ceramic processing temperatures can be incorporated. The impetus for so many combinations of polymers and oxides is, in fact, the sol–gel process. The concept of modifying inorganic networks by using covalent bonding works when the M–C (M is the metal species: aluminum, silicon, titanium, zirconium, etc.) bond is stable and does not undergo hydrolysis under the conditions of the sol–gel process.

1.1 Physical Nanocomposites

(a) Interpenetrating networks

A major classification of hybrids is interpenetrating networks, characterized by a time factor, i.e., sequential or simultaneous. The terminology of interpenetrating polymer networks (IPNs) is borrowed from organic polymer technology (Sperling 1981). In some cases semisequential is used to reflect the linear structure of the organic polymer in which a cross-linked inorganic network is formed.

Some of the earliest inorganic–organic hybrids were prepared by monomer infiltration into a previously formed silica gel. This is an example of a sequential interpenetrating network where the inorganic is the matrix. Specifically, when a dried silica gel (xerogel) is filled with monomer, the monomer can be polymerized *in situ*. Interpenetration is achieved when the impregnating monomer polymerizes in the open pores of the rigid, fully cross-linked silica matrix. In the polymer-impregnated gels, some porosity typically remains. Although some copolymers of methyl methacrylate, butadiene, and styrene have been used as polymers to impregnate silica, the best-known system is silica impregnated with PMMA (Klein 1994). While this type of hybrid was important at first, it has been surpassed by other methods that are simpler.

Another type of sequential interpenetrating network uses silica precursors (TEOS or tetramethylorthosilicate (TMOS)), which are able to solvate some organic polymers (Beaudry and Klein 1996). The polymer solubility enables the silica precursor to polymerize in the environment of an organic polymer solution. This is an example of a sequential interpenetrating network where the organic phase is the matrix. The number of polymers that form solutions with sol–gel formulations is, however, limited. Some initially soluble polymers tend to precipitate during gelation when a change in solvent composition leads to phase separation.

While only a few organic polymers are soluble in sol–gel formulations, many organic monomers are soluble in TEOS (Wojcik and Klein 1995a). These monomers can be introduced directly into sol–gel formulations. Both ring opening polymerizations and free radical polymerizations can be carried out simultaneously with the hydrolysis–condensation of TEOS. To reflect their synchronous polymerizations, the name is simultaneous inorganic–organic interpenetrating polymeric networks. This simultaneous route captures insoluble organic polymers within a sol–gel inorganic network. Generally, the resulting hybrids have no covalent linkages between the organic and inorganic components. The exception occurs when the organic monomer has hydroxyl or carboxyl groups that can lead to M–O–links.

In the preparation of simultaneous networks, it is important to control both polymerizations rates. On the one hand, systems with inorganic condensation rates much faster than the organic polymerization rates turn into brittle hybrids that shrink because the polymer content drops when unreacted monomer evaporates. On the other hand, systems with large organic polymerization rates usually show uncontrolled polymer precipitation. In practice, the kinetics of polymerization are difficult to control, so the success of the simultaneous approach rests on the careful selection of the monomers.

The idea of simultaneous interpenetrating networks has been advanced by Ellsworth and Novak (1993). Several monomers of acrylate type (free radical polymerization) or cyclic alkenes (ring opening metathesis) have been used. They include the acrylamide/*N*,*N'*-bisacrylamide monomer system, 2-hydroxyethyl or hydroxymethyl acrylate, and 7-oxanorbornene and its derivatives. In order to get nonshrinking hybrids, tetraacryloxyethoxysilane has been used instead of TEOS. The acrylate monomer liberated due to hydrolysis acts as a cosolvent to solvate the polymerizing silica network. The result is transparent hybrids with no shrinkage. Wojcik and Klein (1994) have studied the incorporation of silica with di- and triacrylate monomers, hexanediol diacrylate (HDDA), and glyceryl propoxy triacrylate (GPTA). These simultaneous IPNs are transparent hybrids with good mechanical strength.

For functional applications, inorganic gels can be doped with small quantities of organic molecules. Doped gels are prepared either by addition of organics into sol–gel solutions, or by infiltration of the dopant into previously formed xerogels. Although some interactions between the dopant molecules and the inorganic matrix take place, the optical properties or chemical functions of the organics are preserved in the solid state. A number of organics have been used, including organic dyes and chromophores, liquid crystals and active biomolecules, and enzymes and proteins, in order to produce specific optically active or bioactive devices for a broad range of optic and electro-optic applications. Suitably doped gels exhibit highly efficient luminescence, photochemical hole burning, photochromism, and tunable laser action (Avnir *et al.* 1994). In pure inorganic matrices, the activity of the dopant has been found to decrease because of thermodynamic incompatibility with the inorganic phase. Consequently, organic-doped inorganic gels have been replaced with organic-doped hybrid gels. A good example is PMMA–silica composites doped with organic laser dyes, where the hybrid has better properties as a laser host than either pure PMMA or pure silica.

(b) Soluble polymers

For those select polymers that are soluble in metal alkoxide solutions, the preparation of inorganic–organic hybrids is very simple. Taking again the example of silica, the polymerization of the silica precursor occurs around existing polymer chains or domains. Generally it is found that polymerization of TEOS in the presence of polymers under acidic conditions generates small, well-dispersed silica particles. Depending on the choice of polymer, highly transparent, lightweight materials can be obtained. Among the polymers used to form these sequential IPNs, those that give transparent hybrids are those capable of forming hydrogen bonds with hydroxylated silica particles (Landry *et al.* 1992a). Strong interactions between silanols, having the character of Brønsted acids, and specific groups on the polymer that are hydrogen acceptors are responsible for the high degree of two-phase mixing.

Linear polymers with hydrogen acceptor groups, amide, carbonyl, and carbinol, have been used in semisequential inorganic–organic IPNs. These polymers are poly(*N*,*N*-dimethacrylamide), poly(2-methyl-2-oxazoline), PMMA, poly(vinyl pyrrolidone) (PVP), poly(acrylic acid) (PAA) and its copolymers, PVAc, and poly(vinyl alcohol) (PVA) (Landry *et al.* 1992b).

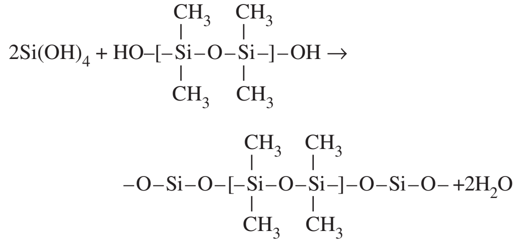
In detail, the properties of the hybrids are functions of the relative fractions of the polymer and TEOS in the sol–gel formulation, along with the chemistry and length of the polymer chain. When TEOS is polymerized in the presence of the polymer, the glass transition temperature (T_g) of the polymer increases or becomes undetectable. The density of the hybrid does not follow a mixing rule for the inorganic and organic components, but is suppressed even with high inorganic contents. More or less porous structures can be formed with open or closed pores. We have studied the optical properties, mechanical properties, and phase separation in sol–gel formulations containing PVAc and poly(ethylene oxide) PEO. These are useful model systems for understanding the behavior in physical nanocomposites (Klein *et al.* 1997).

1.2 Chemical Nanocomposites

(a) Monomer systems

The purely inorganic three-dimensional network in the conventional sol–gel process consists of M–O–M units in the general case, or Si–O–Si units for silica. The purely organic polymer consists of C–C covalent bonds, in the general case, between mers in a polymer backbone, along with weaker van der Waals bonds between polymer chains. However, in a sol–gel process carried out with a precursor containing direct inorganic–organic links, it is possible to form homogeneous materials with C–C, M–C, and M–O bonds. This class of hybrids constitutes inorganic–organic polymers. Undoubtedly, the largest class of true inorganic–organic hybrids is made up of hybrids linked via Si–C bonds that are stable and do not undergo hydrolysis under the sol–gel processing conditions.

There are several methods to link covalently polymers to silica. In an early example, silanol-terminated poly(dimethylsiloxane) (PDMS) was co-condensed with TEOS:



$$2\mathrm{Si(OH)_4} + \mathrm{HO{-}[{-}\underset{CH_3}{\overset{CH_3}{Si}}{-}O{-}\underset{CH_3}{\overset{CH_3}{Si}}{-}]{-}OH} \rightarrow$$

$$\mathrm{{-}O{-}Si{-}O{-}[{-}\underset{CH_3}{\overset{CH_3}{Si}}{-}O{-}\underset{CH_3}{\overset{CH_3}{Si}}{-}]{-}O{-}Si{-}O{-}} + 2\mathrm{H_2O} \qquad (5)$$

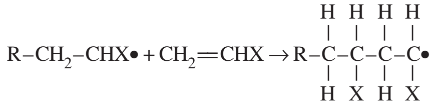
to give inorganic–organic PDMS chains, containing nonbridging methyl groups and nonlinear Si–O–Si angles, which exhibited significant flexibility (Mackenzie *et al.* 1992). When the functionality of PDMS is increased to six (the number of alkoxysilyl groups per oligomeric chain) or PDMS is replaced by poly(tetramethylene oxide) (PTMO), the stiffness and mechanical strength improve considerably (Wang *et al.* 1991).

In a way similar to PDMS and TEOS, other organic polymers have been functionalized with alkoxysilyl groups and covalently bonded to silica. Polymers that have been coupled with silica are PTMO-based polyurethane oligomers, polyoxazolines, polyimide (Mascia and Kioul 1995), poly(arylene ether ketone), poly(arylene ether sulfone), polystyrene, polyoxopropylene (PPO), polyacrylonitrile, copolymers of methyl methacrylate and allyl methacrylate (Wei *et al.* 1990), and cyclophosphazenes (Guglielmi *et al.* 1994).

Naturally, the properties of inorganic–organic hybrids are governed by their microstructure. For example, cellular structures are observed in PDMS–silica gels when the sol–gel parameters (water, HCl concentration, and temperature) are adjusted to accelerate the gelation process. With high concentrations of the organic component, hybrids have low densities, indicating high free volumes characteristic of the organic polymers.

Another method leading to covalent linkage of organic polymer to silica is free radical polymerization. When an organic radical R′ in an organo-substituted silicoester of formula $\mathrm{R'Si(OR)_3}$ contains olefinic moieties, it is possible to build a polymeric network by thermal or photochemical free radical polymerization of C = C bonds, for example:

$$R\bullet + CH_2 = CHX \rightarrow R{-}CH_2{-}CHX\bullet \qquad (6)$$



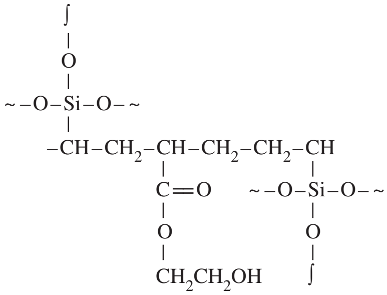
$$R{-}CH_2{-}CHX\bullet + CH_2{=}CHX \rightarrow R{-}\overset{H}{\underset{H}{C}}{-}\overset{H}{\underset{X}{C}}{-}\overset{H}{\underset{H}{C}}{-}\overset{H}{\underset{X}{C}}\bullet \qquad (6a)$$

where R• is a free radical and X is Cl, acetate, CH_2CCl_2, etc. An initiator to produce the free radical is often benzoyl peroxide, $(C_6H_5)_2C_2O_4$. Epoxide groups are also polymer-formers via cationic polymerization or via condensation with curing agents bearing groups such as amine or hydroxyl. Organic polymerization can be carried out simultaneously with condensation of hydrolyzed alkoxysilyl groups.

The commonly used organoalkoxysilanes bearing polymerizable C = C bonds are 3-methacryloxypropyl trimethoxysilane (MEMO) ($CH_2 = C(CH_3)COO(CH_2)Si(OCH_3)_3$) and vinyl trimethoxysilane (VTMS) ($CH_2 = CH{-}Si(OCH_3)_3$) or their ethoxy analog. Alkoxysilyl groups of these inorganic–organic monomers form Si–O–Si bonds upon condensation or co-condensation with another silica precursor such as TEOS. At the same time, (meth)acryl or vinyl groups are able to polymerize with inorganic–organic monomers or with (meth)acrylic or vinyl monomers. These processes lead to the formation of inorganic–organic polymers and copolymers.

Reportedly, inorganic polycondensations of inorganic–organic monomers by the sol–gel process from single monomer systems are difficult. For most of the inorganic–organic precursors, gelation times under acidic conditions are quite long. The reason for the long gelation time is the low concentration of alkoxysilyl groups and steric hindrance from organics in the system. However, the process can be accelerated by incorporation of TEOS.

An idealized structure of an inorganic–organic copolymer synthesis by co-condensation and copolymerization of the tri-monomer system TEOS–vinyltriethoxysilane (VTES)–hydroxyethyl methacrylate (HEMA) is shown below. To accelerate gelation and at the same time reduce shrinkage, inorganic oligomers are used instead of inorganic precursors.



(7)

Ellsworth and Novak (1993) used soluble oligomers of silicic acid instead of TEOS in a condensation with (meth)acryl trialkoxysilane to increase the silica content of inorganic–organic hybrids and reduce shrinkage. Hoebbel *et al.* (1994) used tetramethylammonium silicate ($(N(CH_3)_4)_8Si_8O_{20}\bullet 69H_2O$) converted by an ion-exchange process into double four-ring oligomers of silicic acid ($H_8Si_8O_{20}$). This oligomer can be stabilized by reaction with functional siloxanes containing polymerizable groups like vinyl, allyl, hydrido, or chloromethyl. The functionalized derivative can be polymerized while preserving silicic acid cages in the structure.

When the inorganic–organic monomer bears two or more olefinic moieties in its structure, the monomer becomes highly reactive, creating very favorable conditions for rapid cross-linking by photopolymerization. A series of diacrylate alkoxysilanes was synthesized by Wolter *et al.* (1992) by coupling mercapto-substituted alkoxysilanes to the triacrylate monomers or acrylate resins. The resulting inorganic–organic monomer can be co-condensed via its alkoxysilyl groups with TEOS, whereas its acrylic groups can be cured by UV light, or polymerized thermally. Likewise, diacrylotrialkoxy silane inorganic–organic monomer from aliphatic triacrylate-glycerol propoxy triacrylate (GPTA) and 3-mercaptopropyl trimethoxysilane was used to synthesize UV-curable coatings for optical fibers (Wojcik *et al.* 1993).

A subgroup of hybrids containing a covalent link are those where the polymer exhibits specific electronic or electro-optic properties. Examples of these hybrids include nonlinear optics (NLO) polymer–silica gels, organic dye–silica gels, and conducting polymer–silica gels. For example, organic dye–silica gels of cyan, magenta, and yellow were synthesized at Kodak, where the chromophore was linked covalently to the polymeric component via the reaction of an amine, halogen, or isocyanate functionality, while the silica and polymeric components were covalently bonded via the action of 3-chloropropyltrimethoxysilane. Polymers with conducting properties like polypyrrole, polyaniline, or polythiophene have been covalently incorporated into silica, following their alkoxysilyl functionalization (Sanchez *et al.* 1994).

(b) Monomer–polymer systems

When the organic component is a low molecular weight group or radical and not a polymeric chain or network, the hybrid is an organically modified gel (ormogel) as distinct from an ormosil. Ormogels are formed from $R'Si(OR)_3$-type organosilanes where radical R has a network-modifying effect and is used to introduce new functionalities into the inorganic network. For example, hydrophobicity can be achieved by introducing a methyl, phenyl, or fluoroalkyl chain for R (Schmidt *et al.* 1986). Polymerization of

$R'Si(OR)_3$ alone leads to formation of silsesquioxanes of general formula $(RSiO)_{1.5}$, a silicate framework where each silicon is linked covalently to an organic radical R. A new class of hybrids is emerging called polyhedral oligomeric silsesquioxanes (POSS) that can be blended into rubbers, resins, thermoplastics, and thermosets to enhance thermal stability (Lichtenhan *et al.* 1995).

Bifunctional silanes of formula $R'R''Si(OR)_2$ are used also to prepare ormogels. Upon polymerization, these organoalkoxysilanes form cyclic species or linear chains that are not able to cross-link and form networks themselves. They can be used with tetrafunctional silanes (TEOS) to incorporate desirable functionality into the hybrid network. There are also ormogels where radical R forms a bridge between two silicon atoms. Good examples of these are arylene or alkylene-bridged polysilsesquioxanes synthesized from bis(triethoxysilyl) or bis(trichlorosilyl) aryl or alkyl monomers (Shea *et al.* 1989). Various aliphatic or aromatic molecules have been used as the organic spacer between two silicon atoms.

(c) Copolymers

Copolymers are the products of addition polymerization between two different monomers. Some examples have been mentioned in the sections on physical nanocomposites, as well as chemical nanocomposites. Between two linear polymers such as PVC and PVAc, the units can be arranged in a random copolymer, block copolymer, or graft copolymer. The interesting aspect of copolymers is that the properties of the copolymer are often quite different from the properties of either homopolymer.

In the context of nanocomposites, the term copolymer is applied to polymers with organic and inorganic units. Here especially, the properties of the copolymer are different from the properties of either component. Systems with two or more monomers open opportunities for designing properties, using surface treatments, coupling agents, and sophisticated manufacturing methods.

To recap on chemical nanocomposites, we find that covalently bonded hybrids combine the flexibility and mechanical properties of the organic polymer chains with the hardness and stiffness of the inorganic polymer. In detail, the properties of the hybrids are governed by the chemistry, presence of side chains or pendant groups, and length of polymeric chains. Especially important are the concentration and distribution of alkoxysilyl or equivalent groups, as they directly affect the rigidity or flexibility of the system.

2. *Classes of Polymers in Nanocomposites*

The five classes of polymers that have been incorporated into organic–inorganic hybrids, principally using the sol–gel process, are discussed.

2.1 Elastomers

An elastomer is a polymer with high elasticity due to its coiled structure. A typical elastomer is nonvulcanized rubber, where the chains are first unkinked when a stress is applied, followed by chain straightening and lengthening. Elastomers are poorly crystallized but become more crystalline as they are stretched. Some examples of elastomers include rubbers, polyurethane, and PDMS.

2.2 Partially Crystalline Polymers

Crystallization in polymers is often difficult, leading to glassy or partially crystalline polymers. In general, linear polymers are easier to crystallize than network polymers. Isotactic (the same stereostructure) polymers are easier to crystallize than atactic (random positions) or syndiotactic (alternating positions) polymers. Likewise, it is easier to crystallize polymers in the *trans* (unsaturated positions on opposite sides of the chain) than in the *cis* (unsaturated positions on same side of the chain) configuration. Condensation polymers such as poly(ethylene glycol terephthalate) (Dacron), polyimide, and nylon exhibit some crystallinity because of their moderately stiff chain structures.

2.3 Glassy Polymers

A common example of a glassy polymer is PMMA, where it is difficult to crystallize PMMA because of the atactic structure following addition polymerization. Glassy polymers exhibit a glass transition characterized by the temperature range where the polymer changes its behavior from glassy to rubbery.

2.4 Liquid-Crystalline Polymers

An example of a liquid-crystalline (LC) polymer is hydroxypropyl cellulose (HPC) or 4′-pentyl-4-biphenylcarbonitrile. The characteristic of LC polymers is that they align what are called nematogenic compounds in response to electric fields. LC polymers encapsulated in inorganic materials have been developed for displays. The inorganic encapsulation gives the liquid crystals thermal stability.

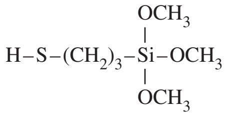
2.5 Thermosets

Thermosetting polymers are polymers that do not soften when heated. Instead, a thermoset polymerizes further with increasing temperature, resulting in a continuous network where polymer units cannot slip past each other. An example of a thermosetting polymer is phenol formaldehyde, where phenol (C_6H_6O) and formaldehyde (CH_2O) undergo condensation polymerization to generate phenol formaldehyde that is able to link at a minimum of three places and a maximum of five places.

3. *Weak and Strong Bonding in Nanocomposites*

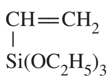
3.1 Covalently Bonded Systems

The majority of hybrids investigated have covalent bonds between the organic and inorganic components. These bonds are formed before incorporation into a nanocomposite. A typical route involves a metal alkoxide that has been functionalized via a mercapto group (Wojcik and Klein 1995b):



$$H{-}S{-}(CH_2)_3{-}\overset{\displaystyle OCH_3}{\underset{\displaystyle OCH_3}{Si}}{-}OCH_3 \qquad (8)$$

which is used during the creation of the nanocomposite when a reactive site, such as a vinyl group:



$$\underset{\displaystyle Si(OC_2H_5)_3}{CH}{=}CH_2 \qquad (9)$$

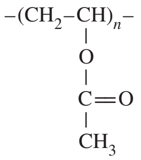
is available and covalently bonded to the silicon atom. The reason that forming direct linkages has been popular is, in large part, that it ensures strong bonds between the components.

3.2 Hydrogen-Bonded Systems

The minority of hybrids investigated have hydrogen bonds between the organic and inorganic components. These bonds are formed between hydroxyls on the inorganic gel and ether oxygens on the polymers assuming a resonant hybrid pair for PEO:

$$[+CH_2-CH_2-O-/+O-CH_2-CH_2-]\rightarrow$$
$$-(O-CH_2-CH_2-)_n- \qquad (10)$$

or carbonyl groups on the polymers:



$$-(CH_2-\underset{\displaystyle \underset{\displaystyle \underset{\displaystyle CH_3}{C=O}}{O}}{CH})_n- \qquad (11)$$

as in the case of PVAc. The reason that there are few hydrogen-bonded hybrids is directly related to the small number of water-soluble monomers and the even smaller number of soluble polymers. Nevertheless, the technological importance of the hydrogen-bonded systems is evident in the development of organically modified electrolytes (ormolytes) (Charbouillot *et al.* 1988).

3.3 Phase Separation: Phenomena and Kinetics

As described above, physical hybrids depend on the polycondensation of an inorganic network in the presence of an organic polymer dissolved in the solution, or polymerization of organic monomers inside the porous inorganic network. To obtain homogeneous mixing, the degree of phase separation has to be controlled. There should be strong, attractive interactions between the organic and inorganic polymers to achieve good interfacial bonding. Polymers containing carbonyl groups, such as PVAc, can be incorporated into sol–gel processes because they are soluble in alcohol–water mixtures due to partial esterification of the side chains. Polymers containing ether oxygens, such as PEO, sometimes referred to as poly(ethylene glycol) (PEG), are soluble in water because the ether oxygens readily form hydrogen bonds.

Landry *et al.* (1992b) have demonstrated the simplicity of the approach with TEOS and a number of polymers. Depending on (a) the synthesis conditions and (b) the relative amounts of organic and inorganic polymer, the morphologies ranged from discrete silica particles within an organic matrix, to interpenetrating networks, to polymeric domains dispersed within a silica matrix. Under acidic conditions with TEOS it was possible to form optically transparent films with PMMA, PVAc, poly(vinyl pyrrolidone), and poly(*N*,*N*-dimethacrylamide). Hydrogen bonding between the silanols on the silica polymer and the carbonyl groups on the polymers prevented macroscopic phase separation. Phase separation, if any, occurred on such a small scale that transparent materials were formed without the need for covalent bonding. Furthermore, there is no evidence for chemical reactions below 150 °C. Acetic acid was evolved from PVAc above 200 °C, but this is likely to be a decomposition product and not an indication of covalent bonding between TEOS and PVAc.

Polymerization of TEOS in the presence of a soluble polymer does not always lead to a homogeneous mixture. Kaji *et al.* (1993) have taken advantage of the incompatibility of PAA and silica and PEO and silica. Their studies have shown that the scale of the phase separation and its timing are dependent on (i) the amount of polymer, (ii) the molecular weight of the polymer, (iii) the solvent composition, and (iv) the pH. The gel morphologies are the results of a competition between phase separation and gelation, also referred to as "chemical quenching." The rate of gelation of the silica network is in opposition to the rate of phase separation, which is an indication of the segregation strength. When gelation is faster than phase separation, the system becomes rigid with a morphology reflecting the early structure of the gel. When phase separation is faster than gelation, the system becomes rigid with a variety of transient structures that are increasingly heterogeneous, with the extreme case being sedimentation.

4. Importance of Interfaces in Inorganic–Organic Nanocomposites

The interface between the organic and the inorganic constituents is the key issue in hybrids and nanocomposites. In our work we are measuring strengths, hardness, and skeletal densities of some representative systems. Mechanical behavior is at best an indirect measurement of interface behavior. However, by relating these properties to bonding and microstructure, we try to examine the nature of the interface and the scale on which it controls properties.

Other properties, such as optical, electronic, magnetic, catalytic, and chemical durability require further investigation. We need to understand what these properties tell us about the interfaces, so that we can better engineer the interfaces.

Finally, we should address the questions, what analytical techniques are available to probe the interface? Are indirect measurements more or less meaningful than direct measurements, considering that performance may be a reflection of averaged properties? Other articles in this encyclopedia should shed some light on these subjects.

Bibliography

Andrews M P, Najafi S I (eds.) 1997 *Critical Reviews: Sol–Gel and Polymer Photonic Devices* (CR68-03). SPIE, Bellingham, WA

Avnir D, Braun S, Lev O, Levy D, Ottolenghi M 1994 Organically doped sol–gel porous glasses: chemical sensors, enzymatic sensors, electrooptical materials, luminescent materials and photochromic materials. In: Klein L C (ed.) *Sol–Gel Optics*. Kluwer, Boston, pp. 539–82

Avnir D, Klein L C, Levy D, Schubert U, Wojcik A B 1998 Organo-silica sol–gel materials. In: Rappoport Z, Apeloig Y (eds.) *The Chemistry of Organosilicon Compounds*. Wiley, London, Vol. 2, pp. 2317–62

Beaudry C L, Klein L C 1996 Sol–gel processing of silica/poly(vinyl acetate) (PVAc) nanocomposites. In: Chow G-M, Gonsalves K E (eds.) *Nanotechnology: Molecularly Designed Materials*. American Chemical Society, Washington, DC, pp. 382–94

Brinker C J, Scherer G W 1990 *Sol–Gel Science*. Academic Press, Boston

Charbouillot Y, Ravaine D, Armand M, Poinsignon C 1988 Aminosils: new solid state protonic materials by the sol–gel process. *J. Non-Cryst. Solids* **103**, 325–30

Chujo Y Saegusa T 1992 Organic polymer hybrids with silica gel by means of the sol–gel method. In: Advances in Polymer Science. Springer, Berlin, Vol. 100, pp. 12–29

Coltrain B K, Sanchez C, Schaefer D W, Wilkes G L (eds.) 1996 *Symp. Proc. 435, Better Ceramics Through Chemistry VII: Organic/Inorganic Hybrid Materials*. Materials Research Society, Pittsburgh, PA

Drzal L T, Opila R L, Peppas N A, Schutte C (eds.) 1995 *Symp. Proc. 385, Polymer/Inorganic Interfaces II*. Materials Research Society, Pittsburgh, PA

Ellsworth M W, Novak B M 1993 "Inverse" organic–inorganic composite materials: 3. High glass content "nonshrinking" sol–gel composites via poly(silicic acid esters). *Chem. Mater.* **5**, 839–44

Guglielmi M, Colombo P, Brusetin G, Facchin G, Gleria M 1994 New materials based on the reaction of cyclo- and poly-(organo phosphazenes) with SiO_2, TiO_2 and ZrO_2 precursors. *J. Sol–Gel Sci. Technol.* **2**, 109–14

Hoebbel D, Endres K, Reinert T, Pitsch I 1994 Inorganic–organic polymers derived from functional silicic acid derivatives by additive reaction. *J. Non-Cryst. Solids* **176**, 179–85

Huang H H, Orler B, Wilkes G L 1985 Ceramers: hybrid materials incorporating polymeric/oligomeric species with inorganic glasses by a sol–gel process: 2. Effect of acid content on the final properties. *Polym. Bull.* **14**, 557–64

J. Sol–Gel Sci. Technol. 1995 **5**(2), 75–157

J. Sol–Gel Sci. Technol. 1996 **7**(3), 151–224

Kaji H, Nakanishi K, Soga N 1993 Polymerization-induced phase separation in silica sol–gel systems containing formamide. *J. Sol–Gel Sci. Technol.* **1**, 35–46

Klein L C 1994 Nanocomposite fabrication for transparent windows. In: Klein L C (ed.) *Sol–Gel Optics*. Kluwer, Boston, pp. 215–32

Klein L C, Beaudry C L, Yamazaki S, Wojcik A B 1997 Transparent polymer/silica hybrid gels. In: Andrews M P, Najafi S I (eds.) *Sol–Gel and Polymer Photonic Devices*. SPIE, Bellingham, WA, pp. 54–63

Klein L C, Francis L F, DeGuire M R, Mark J E (eds.) 1999 *Symp. Proc. 576, Organic/Inorganic Hybrid Materials II*. Materials Research Society, Pittsburgh, PA

Laine R M, Sanchez C, Brinker C J, Giannelis E (eds.) 1998 *Symp. Proc. 519, Organic/Inorganic Hybrid Materials*. Materials Research Society, Pittsburgh, PA

Landry C J T, Coltrain B K, Brady B K 1992b In situ polymerization of tetraethoxysilane in poly(methyl methacrylate). *Polymer* **33**, 1486–95

Landry C J T, Coltrain B K, Wesson J A, Zumbulyadis N, Lippert J C 1992a In situ polymerization of tetraethoxysilane in polymers. *Polymer* **33**, 1496–506

Lerner M M, Oriakhi C O 1997 Polymers in ordered nanocomposites. In: Goldstein A N (ed.) *Handbook of Nanophase Materials*. Marcel Dekker, New York, pp. 199–219

Lichtenhan J D, Otonari Y A, Carr M J 1995 Linear hybrid polymer building blocks: methacrylate-functionalized polyhedral oligomeric silsesquioxane monomers and polymers. *Macromolecules* **28**, 8435–7

Mackenzie J D, Chung Y J, Hu Y 1992 Rubbery ormosils and their applications. *J. Non-Cryst. Solids* **147/148**, 271–9

Mascia L, Kioul A 1995 Influence of siloxane composition and morphology on properties of polyimide–silica hybrids. *Polymer* **19**, 3649–59

Sanchez C, Alonso B, Chapusot F, Ribot F, Audebert P 1994 Molecular design of hybrid organic–inorganic materials with electronic properties. *J. Sol–Gel Sci. Technol.* **2**, 161–6

Sanchez C, Ribot F 1994 Design of hybrid organic–inorganic materials synthesized via sol–gel chemistry. *New J. Chem.* **18**, 1007–47

Schmidt H 1994 Inorganic–organic composites by sol–gel techniques. *J. Sol–Gel Sci. Technol.* **1**, 217–31

Schmidt H, Scholze H, Tunker G 1986 Hot melt adhesives for glass containers by the sol–gel process. *J. Non-Cryst. Solids* **80**, 557–63

Schmidt H, Wolter H 1990 Organically modified ceramics and their applications. *J. Non-Cryst. Solids* **121**, 428–35

Shea K J, Loy D A, Webster O W 1989 Aryl-bridged polysilsesquioxanes—new microporous materials. *Chem. Mater.* **1**, 572–4

Sperling L H 1981 *Interpenetrating Polymer Networks and Related Materials*. Plenum, New York

Wang B, Wilkes G L, Hedrick J C, Liptak S C, McGrath J E 1991 New high refractive index organic/inorganic hybrid materials from sol–gel processing. *Macromolecules* **24**, 3449–53

Wei Y, Bakthavatchalam R, Whitecar C K 1990 Synthesis of new organic–inorganic hybrid glasses. *Chem. Mater.* **2**, 337–9

Wen J, Wilkes G L 1996 Organic/inorganic hybrid network materials by the sol–gel approach. *Chem. Mater.* **8**, 1667–81

Wojcik A B, Klein L C 1994 Organic–inorganic gels based on silica and multifunctional acrylates. *J. Sol–Gel Sci. Technol.* **2**, 115–20

Wojcik A B, Klein L C 1995a Transparent inorganic/organic copolymers by the sol–gel process: thermal behavior of copolymers of tetraethyl orthosilicate (TEOS), vinyl triethoxysilane (VTES) and (meth) acrylate monomers. *J. Sol–Gel Sci. Technol.* **5**, 77–82

Wojcik A B, Klein L C 1995b Transparent inorganic/organic copolymers by the sol–gel process: copolymers of tetraethyl orthosilicate (TEOS), vinyl triethoxysilane (VTES) and (meth)acrylate monomers. *J. Sol–Gel Sci. Technol.* **4**, 57–66

Wojcik A B, Klein L C 1997 Organic/inorganic hybrids by the sol–gel process: a classification scheme. *Appl. Organomet. Chem.* **11**, 129–35

Wojcik A B, Klein L C, Rondinella V V, Matthewson M J, Foy P 1993 Organically modified silicate coatings for optical fibers. SPIE 2074, pp. 135–41

Wolter H, Glaubitt W, Rose K 1992 Multifunctional (meth)acrylate alkoxysilanes: a new type of reactive compounds. In: Hampden-Smith M J, Klemperer W G, Brinker C J (eds.) *Symp. Proc. 271, Better Ceramics Through Chemistry V*. Materials Research Society, Pittsburgh, PA, pp. 719–24

L. C. Klein
Rutgers University, Piscataway, New Jersey, USA

A. B. Wojcik
Hybrid Glass Technologies, Princeton, New Jersey, USA

Polymeric Discotic Liquid Crystals

Discotic mesogens composed of a prolate and often rigid core, to which flexible alkyl side chains are attached can show liquid crystalline (LC) behavior (Demus *et al.* 1998). Their LC phase behavior can range from a loose parallel orientation of the mesogens with no positional order (like a random pile of coins), called the nematic discotic N_D phase, to a highly ordered hexagonal packing of columns (like a regularly packed stack of coins), termed the ordered hexagonal columnar phase Col_{ho} (formerly known as discotic hexagonal ordered D_{ho} phase). Depending on the mesogen packing inside the column (ordered, disordered, or tilted) and the arrangement of the columns in spatial lattices (hexagonal, rectangular, or oblique), a variety of different phase structures are observed.

When the discotic mesogens are interconnected by linking units, polymers are obtained, which often retain or even stabilize the mesophases of the monomers (Müllen and Wegner 1998). Figure 1 shows the different types of discotic polymer structures which can be divided into the three general classes of rigid rod polymers, semi-flexible polymers, and networks (Ringsdorf *et al.* 1991, pp. 211–71). The rigid rod structure can be obtained by either forming the stiff polymer backbone along the central axis of the columnar arrangement (Fig. 1(a)) or by linking the discotic mesogens with stiff spacers on the periphery of the mesogens (Fig. 1(b)). For the semi-flexible polymers two different types are known: the main chain structure with the mesogens interconnected by flexible spacers (Fig. 1(c), like beads on a thread), or the side group polymer (SGP) with the mesogens attached to the flexible polymer backbone by spacer units (Fig. 1(d)). A special case of SGPs with a flexible backbone are the phasmidic systems (Fig. 1(e)), which form columnar mesophases. In these materials the mesogens are flat, tapered units (like pieces of cake) that self-assemble into columns (formerly Φ, now Col). If the mesogens carry more than one polymerizable unit, cross-links between backbones can be formed which leads to insoluble networks and elastomers (Fig. 1(f)).

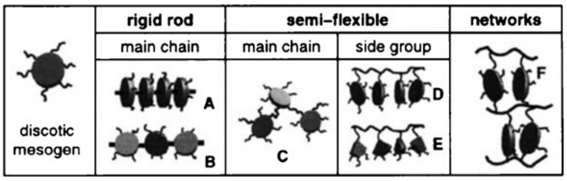


Figure 1
Schematic structures of polymeric discotic liquid crystals.

Most of the known discotic polymers are thermotropic liquid crystals, only very few examples of lyotropic discotic polymers have been described in the literature.

1. *Mesogens and Polymer Structures*

Since their discovery by Chandrasekhar in 1977, a vast number of different discotic mesogens have been synthesized, and their LC phase behavior has been studied (Chandrasekhar 1993). From this large variety only a relatively small number of mesogens were used to synthesize polymeric analogs. Selected examples are shown in Fig. 2 subdivided by their general polymer structure.

The spinal rigid rod polymers (Fig. 1(a)), composed of cofacially interconnected discotic subunits, are dominated by phthalocyanine **1** mesogens with a metal ion complexed in the central cavity. The metal ion links two adjacent disks by an oxygen bridge which leads to covalent fixation of the columnar

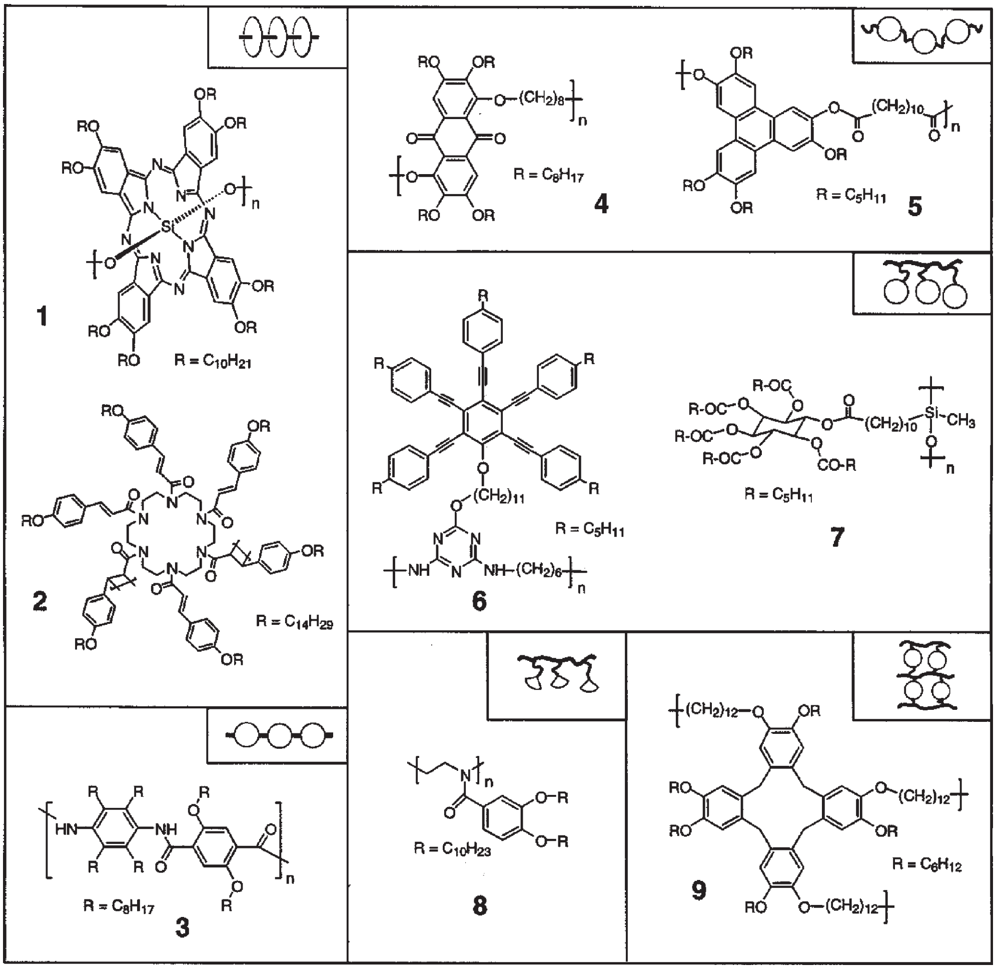


Figure 2
Selected examples of mesogen structures for the different polymer classes: **1** phthalocyaninatopolysiloxane, **2** cinnamoyl-substituted hexaazacrown, **3** sanidic polyamide, **4** rufigallol polyether, **5** triphenylene polyester, **6** penta-alkynylbenzene polyamine, **7** inositol polysiloxane, **8** benzylated polyethylenimine, and **9** cyclotetraveratrylene network.

structure. Tin, silicon, and germanium were used as metals to form these highly shape-anisotropic polymers (Oriol and Serrano 1995). Other examples of stiff polymeric columnar structures are hexaazacrowns **2** and glucopyranoses with cinnamoyl substituents which can be photopolymerized in the columnar mesophase. Here, the interconnections between the disks of one column are located at the periphery of the mesogens (Mukkamala *et al.* 1996).

Rigid rod polymers, which are obtained when the discotic mesogens are linked on their periphery with stiff spacer units (Fig. 1(b)), and which form columnar mesophases, have not yet been identified. However, there are examples of fully aromatic polyesters, polyamides **3** (Fig. 2), and polyazomethines with flexible alkyl sidechains (Voigt-Martin *et al.* 1995) which are derived from discotic monomers, but these polymers show calamitic phase behavior.

In the class of discotic flexible main chain polymers (MCPs, Fig. 1(c)) triphenylenes **5** are the prevalent mesogenic unit. The triphenylene cores are interconnected by flexible alkyl spacers, which allow a columnar packing of the polymeric mesogens similar to the mesophase organization of the low molar mass derivative (Ringsdorf *et al.* 1991 pp. 211–71). Further examples of this polymer class include rufigallol **4** (anthraquinone), tetrasubstituted hydroquinone, and phthalocyanine moieties as discotic groups.

Flexible SGPs (Fig. 1(d) and (e)) are the most abundant structure type in the family of discotic LC polymers. Here, the discotic side groups are dominated again by triphenylene units, but phthalocyanine, penta-alkynylbenzene **6**, cellobiose, and inositol **7** derivatives were also used as mesogens. In this polymer type the discotic subunits have the largest degree of freedom to form columnar phases while the flexible alkyl spacers play a delicate role in decoupling the columns from the polymer backbone (Werth and Spiess 1993). The phasmidic SGPs (Fig. 1(e)) can be seen as a derivative of the corresponding discotic polymers. The columnar mesophase-forming units are fragments of a hypothetical circular discogen which self-assemble into a cylindrical packing (Percec *et al.* 1998). The common structure of the phasmidic mesogens is a tapered benzoic acid derivative with one to three alkyloxy side chains and dendritic structures based upon the benzoic acid subunit **8**.

2. Synthesis

Discotic polymers can be synthesized by three major strategies, partly depending on the target structure of the polymer. The most common procedure (Fig. 3, route I) starts from a discotic mesogen with a functional group that can be derivatized with a polymerizable unit. The obtained discotic monomers are often polymerized in the mesophase to capture the columnar structure. Suitable monomers are acrylates and methacrylates that undergo radical polymerization upon thermal or photochemical activation with an appropriate radical initiator. Some unsaturated systems (like stilbenes or cinnamic acid derivatives) can also undergo photoaddition reactions to form a polymer from the properly arranged monomers. If the discotic monomers carry two or more functional groups, MCPs can be obtained by polycondensation with reactive spacer units. Ring-opening metathesis polymerization (ROMP) has been used for strained cyclic monomers with a double bond in the ring (Weck *et al.* 1997). The unsaturated polymer backbones obtained in this way can be converted into the saturated hydrocarbon by catalytic hydrogenation.

Grafting functionalized discogens onto a preformed polymer which bears reactive site, leads to SGPs (Fig. 3, route II). In this way polysiloxanes with triphenylene side groups could be synthesized by hydrosilylation which represents one of the first examples of discotic LC polymers (Kreuder and Ringsdorf 1983). Metalphthalocyanines were used in a polymer-analogous Friedel–Crafts acylation of phenyl groups in a polyamide to yield lyotropic SGPs. The acylation of linear polyethylenimine with alkoxysubstituted benzoic acid derivatives led to phasmidic SGPs **8** (Fig. 2).

The third route to discotic polymers is based on the cyclization reaction of bifunctional mesogen subunits (Fig. 3, route III). The polymer is formed simultaneously with the discotic core when two mesogen fragments, interconnected by a spacer, react with an excess of single mesogen fragments. Depending on the number of fragments needed to form the discotic mesogen, linear MCPs (two fragments) or networks (⩾three fragments) are formed. The degree of cross-linking in a network is controlled by the relative concentration of mesogen fragment dimer with respect to the single fragment. Networks of polymeric cyclotetraveratrylenes **9** with mesogenic properties could be obtained by this approach.

3. Characterization of Compounds and Phases

In addition to the standard characterization of the chemical structure of newly synthesized discotic LC polymers, special techniques are required to investigate the polymer characteristics and temperature-dependent morphology.

Size exclusion chromatography or gel permeation chromatography (SEC or GPC) has been used to determine the effect of molecular weights on the clearing temperature of discotic LC polymers. As

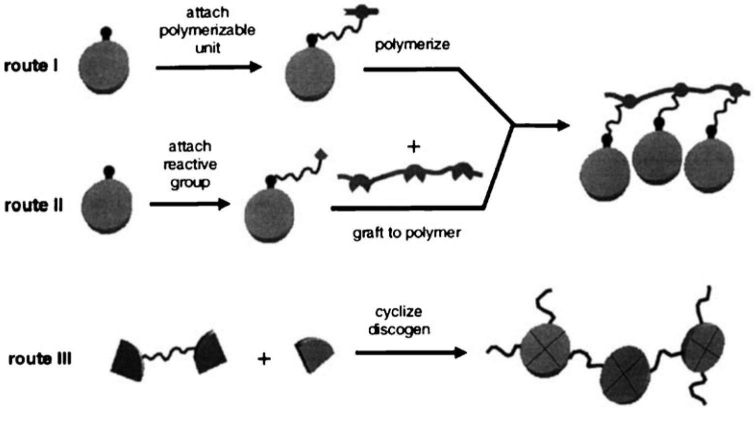


Figure 3
General strategies for the synthesis of discotic polymers.

general trend an increase in clearing temperature is seen with higher degrees of polymerization, reaching a plateau at around 10 monomer units in triphenylene-based MCPs.

To determine if a polymer shows LC behavior, differential scanning calorimetry (DSC) is commonly applied. For monomers the transition from the crystalline state into the mesophase is usually detected by an endothermic peak upon heating. The reverse crystallization process often occurs at a lower temperature, with the hysteresis caused by the high viscosity of the mesophase. In discotic polymers crystallization normally does not take place, but an endothermic glass transition can sometimes be detected upon heating when the polymer undergoes a transformation from the glass to the mesophase. In both the monomer and the polymer the melting points are commonly visible as distinct exothermic peaks.

The transition entropies, ΔS, can be derived by dividing the enthalpies, ΔH, measured with DSC, by the transition temperature ($\Delta S = \Delta H / T_{\mathrm{tr}}$). Marked differences between the transition entropies of homologous derivatives indicate different degrees of order in their mesophases. This was shown for a series of phthalocyanine containing polymers which had the same general structure but with different metal ions complexed in the discotic core.

Temperature-dependent polarizing optical microscopy is often employed to support the result obtained by DSC. The ordered domains formed in the mesophase usually cause characteristic birefringence textures specific to the corresponding mesophase structure. Columnar discotic phases can be identified by the fan-like or digitate star-shaped texture that forms when cooling from the isotropic melt. At sufficiently low viscosity the birefringent pattern can be deformed upon shearing which is a clear indication of the existence of a LC phase. In discotic polymers the viscosity often increases with decreasing temperature so that the mesophase appears frozen at room temperature even above the glass transition.

X-ray diffraction measurements of LC polymers allow the determination of structural parameters, such as columnar distances, packing geometry (e.g., hexagonal, oblique, etc.), stacking distance of mesogenic units within a column, and average interatomic distance between alkyl side chains. The packing distances can be obtained from samples with an isotropic distribution of the mesophase domains (powder diffraction), while the sample must be macroscopically oriented to obtain more detailed information about columnar orientation and position of mesogenic units. Macroscopic alignment can be achieved for polymers in the mesophase by applying external forces such as strain, or by means of magnetic fields, or electrical fields (Fig. 4). The resulting orientation depends on the polymer structure and the nature of the applied force (Hüser *et al.* 1989). In magnetic fields aromatic mesogens (like triphenylenes) turn their flat discotic core parallel to the field vector, due to the diamagnetic character of their conjugated π-systems. As a result, the column axes are oriented perpendicular to the magnetic field vector, but the columnar directors are isotropically distributed within a plane perpendicular to the field. By mechanically drawing a sample in the mesophase, the columns can be oriented macroscopically along a single director. The columnar director orientation with respect to the stress vector depends specifically on the polymer structure since the applied stress primarily affects the polymer backbone rather than the discotic mesogen. Thus, discotic MCPs can be oriented with their backbone parallel and the columnar axes perpendicular to the strain direction. SGPs also orient their backbones in the strain direction. There, however, the discotic columns also align parallel to the strain direction, as observed by x ray scattering experiments. Shear-induced orientation of columnar phthalocyaninatopolysiloxanes upon transfer from the air–water interface by the Langmuir–Blodgett technique is a further example of possible large-scale orientation by mechanical forces.

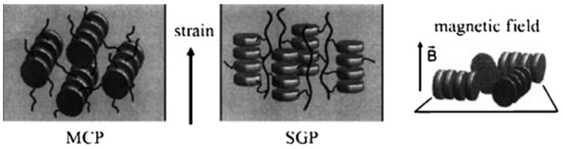


Figure 4
Schematic orientation of the columnar axes in MCPs and SGPs with respect to the strain direction and columnar orientation in the magnetic field.

Transmission electron microscopy (TEM) combines the advantage of direct imaging at almost molecular resolution with the possibility of using electron diffraction for bulk structure determination in ordered materials. This technique was used to study a large series of sanidic polymers to understand the effect of structural modifications on the thermotropic behavior. The direct images showed the layer structure in the crystalline phase and allowed the measurement of single layer thicknesses (Voigt-Martin *et al.* 1995). The electron diffraction patterns at different temperatures were used to monitor structural change in the different phases and enabled the discrimination between LC phases and microcrystallites embedded in an amorphous matrix.

Nuclear magnetic resonance (NMR) spectroscopy is a very powerful tool for resolving the chemical structure of a molecule as well as for measuring dynamic processes of different structural units (Spiess 1993). With this technique free mesogen rotation about the column axis could be observed in the LC phase of monomeric discotic materials. This rotation is hindered in polymeric discotics due to the fixation of the mesogen at the polymer backbone, and only partial movement around ill-defined angles can be observed.

In deuterated samples the different mobilities of the core and the alkyl chains were elucidated by solid state ^{2}H-NMR.

4. Applications

The special anisotropic mechanical, optical, and electronic properties of columnar discotic liquid crystals have stimulated a number of ideas for possible applications.

In calamitic liquid crystals the anisotropic optical characteristics of ordered mesophases have long been utilized for display applications. Low molecular weight discotic liquid crystals have been proposed for optical compensation layers to improve the viewing angle of twisted nematic LC displays (Favre-Nicolin *et al*. 1996). Polymeric discotic materials would combine these useful optical properties with an increased thermal and mechanical stability of the required mesophase and the possibility of macroscopic alignment.

The most appealing property of polymeric discotics is their anisotropic conductivity. In monomeric triphenylenes very high charge carrier mobilities of up to $0.1\,\mathrm{cm^2V^{-1}s^{-1}}$ were found in the columnar mesophase. Such high values motivated the research towards their application as photoconductors in electrophotography, as photovoltaics, and as hole transport layers in organic light emitting diodes (LEDs). Functional devices of such organic LEDs have been reported for discotic MCPs and SGPs (Christ *et al*. 1997, Bacher *et al*. 1997). The change in electrical conductivity of crown ether substituted phthalocyanines upon exposure to reducing and oxidizing gases has been used to demonstrate their possible application as gas sensor material. Spinal polysiloxanes of such crown ether phthalocyanines are also discussed as molecular wires since the peripheral crown ethers stack in a tubular fashion to complex and transport alkali metal ions along the columns (van Nostrum 1996).

5. Concluding Remarks

Low molecular weight discotic liquid crystals are known to have many interesting mechanical, optical, and electronic properties that originate from the highly anisotropic morphology of these materials. For many of the known discotic mesogens polymeric derivatives were synthesized with the aim of combining their LC properties with the useful characteristics of polymers. Still, the number of examples for discotic MCPs, SGPs, and 3D networks are scarce compared to polymeric calamitic systems, which have already found ample technical applications.

The specific advantages of polymeric discotic liquid crystals over their low molecular weight counterparts can be seen in the improved processability and mesophase stability partly resulting from their increased viscosity. The polymeric character also allows the macroscopic alignment of the columnar structures by mechanical stress or magnetic fields which makes these highly anisotropic materials interesting candidates for new optical and electronic applications.

Bibliography

Bacher A, Bleyl I, Erdelen C H, Haarer D, Paulus W, Schmidt H-W 1997 Low molecular weight and polymeric triphenylenes as hole transport materials in organic two-layer LEDs. *Adv. Mater*. **9**, 1031–5

Chandrasekhar S 1993 Discotic liquid crystals. A brief review. *Liq. Cryst*. **14**, 3–14

Christ T, Stümpflen V, Wendorff J H 1997 Light emitting diodes based on a discotic main chain polymer. *Macromol. Rapid. Commun*. **18**, 93–8

Demus D, Goodby J, Gray G W, Spiess H-W, Vill V (eds.) 1998 *Handbook of Liquid Crystals*. Wiley-VCH, Weinheim, Germany

Favre-Nicolin C D, Lub J, van der Sluis P 1996 Optical and structural properties of new discotic acrylates polymerized in the discotic nematic phase. *Adv. Mater*. **8**, 1005–8

Hüser B, Pakula T, Spiess H W 1989 Macroscopic ordering of liquid-crystalline polymers with discotic mesogens. *Macromolecules* **22**, 1960–3

Kreuder W, Ringsdorf H 1983 Liquid crystalline polymers with disc-like mesogens. *Makromol. Chem. Rapid Commun*. **4**, 807–15

Mukkamala R, Burns C L, Catchings III R M, Weiss R G 1996 Photopolymerization of carbohydrate-based discotic mesogens. *J. Am. Chem. Soc*. **118**, 9498–508

Müllen K, Wegner G (eds.) 1998 *Electronic Materials: The Oligomer Approach*. Wiley-VCH, Weinheim, Germany

Oriol L, Serrano J L 1995 Metallomesogenic polymers. *Adv. Mater*. **7**, 348–69

Percec V, Ahn C-H, Ungar G, Yeardley D J P, Möller M, Sheiko S S 1998 Controlling polymer shape through the self-assembly of dendritic side-groups. *Nature* **391**, 161–4

Ringsdorf H, Voigt-Martin I, Wendorff J, Wüstefeld R, Zentel R 1991 Molecular engineering of liquid crystalline polymers. In: Fischer E W, Schulz R C, Sillescu H (eds.) *Chemistry and Physics of Macromolecules*. VCH, Weinheim, Germany, pp. 211–71

Spiess H W 1993 Structure and dynamics of liquid-crystalline polymers with different molecular architectures from multidimensional NMR. *Ber. Bunsenges. Phys. Chem*. **97**, 1294–305

van Nostrum C F 1996 Self-assembled wires and channels. *Adv. Mater*. **8**, 1027–30

Voigt-Martin I, Simon P, Bauer S, Ringsdorf H 1995 Structure and defects in sanidic liquid crystalline polymers. 1. *Macromolecules* **28**, 236–42

Weck M, Mohr B, Maughon B R, Grubbs R H 1997 Synthesis of discotic columnar side-chain liquid crystalline polymers by ring-opening metathesis polymerization (ROMP). *Macromolecules* **30**, 6430–7

Werth M, Spiess H W 1993 The role of the spacer in discotic polymers. *Makromol. Chem. Rapid Commun*. **14**, 239–8

U. Jonas and H. W. Spiess
Max-Planck-Institut für Polymerforschung, Postfach,
Mainz, Germany

Porous Ceramics including Fibrous Insulation, Structure and Properties of

The presence of the porosity in a material is often viewed as problematic. There are, however, many applications in which the use of porous materials can be advantageous, e.g., refractories, filters, membranes, catalytic substrates, thermal insulation, gas burner materials, etc. In these applications, materials are chosen as a result of special properties, such as low thermal conductivity, high permeability, high-temperature stability, low dielectric constants, etc. Although the primary function of these materials may not be structural, many of these applications require a high degree of mechanical reliability. The simplest classification of porous ceramics is with respect to the volume fraction (P) of the porosity. The emphasis here is restricted to high-porosity ceramics, which are defined somewhat arbitrarily as ceramics with $P > 0.70$. These materials are commercially available in two broad groups, fibrous and cellular. Gibson and Ashby (1997) have written a comprehensive review of the literature on cellular materials and Brezny and Green (1994) have reviewed work on cellular ceramic materials. There are two other related groups of high-porosity ceramic materials that are worthy of mention. The first of these groups is based on the consolidation of hollow spheres. For example, some glasses are available in the form of hollow spheres and these spheres can be bonded or sintered together to produce a high-porosity material. The second group is those that are formed by a sol–gel process. When dried, aerogels or xerogels can possess "nanometer-sized" porosity with high levels of porosity.

1. Structure of High-porosity Ceramics

1.1 Fibrous Ceramics

Fibers in a blanket or felt form have been used extensively as thermal insulation but fibers can also be "rigidized" by bonding processes to produce fibrous ceramic monoliths. An example of a fibrous ceramic is shown in Fig. 1. This particular material is based on a bonded network of high-purity silica glass fibers and is used as insulation in the thermal protection system of the NASA space shuttles. In this application, there has also been an interest in using other fiber types (e.g., aluminosilicate fibers) to produce a composite insulation with improved mechanical reliability. Similar materials are used for a wide variety of thermal insulation applications and are attractive as a result of their low thermal mass, which allows fuel savings. These materials are usually classified in terms of the refractoriness of the fibers. They may range from low- to high-purity silica glasses or from low- to high-alumina fibers, depending on the alumina/silica ratio in the fibers. The low-alumina fibers are usually produced by a melt blowing process. High-alumina fibers (ICI Industries, Saffil alumina; or 3M Corporation, Nextel mullite) are produced by a sol–gel route and can be used at temperatures up to 1600 °C. Fibrous ceramic monoliths are often formed by a pressing operation and this can lead to significant anisotropy in the physical properties. The main concern regarding the use of these materials is the classification of some ceramic fibers as possible carcinogens.

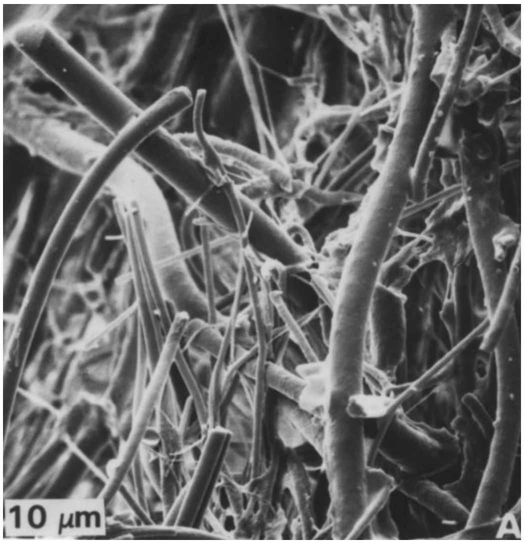


Figure 1
Structure of a "rigidized" fibrous ceramic body. The material shown is used as thermal insulation for the NASA space shuttles.

1.2 Cellular Ceramics

Cellular ceramics are comprised of various arrangements of space-filling polygons (cells) and can be classified into two broad groups, honeycombs and foams. In honeycombs (see Fig. 2), the cells form a two-dimensional array, whereas foams are comprised of a three-dimensional array of hollow polygons. Honeycombs have been used extensively as catalytic substrates in automobile exhaust systems. Foams are usually subdivided into two further categories depending on whether or not the individual cells possess solid faces. If the solid of which the foam is made is contained only in cell edges, the material is termed open cell, i.e., the void space is connected through the cell faces and the material is permeable. If the cell faces are present, the foam is termed closed cell and the individual cells are isolated from each other. There is clearly the possibility that foams can be partly open and partly closed. Figure 3 compares the structures of an open- and closed-cell ceramic.

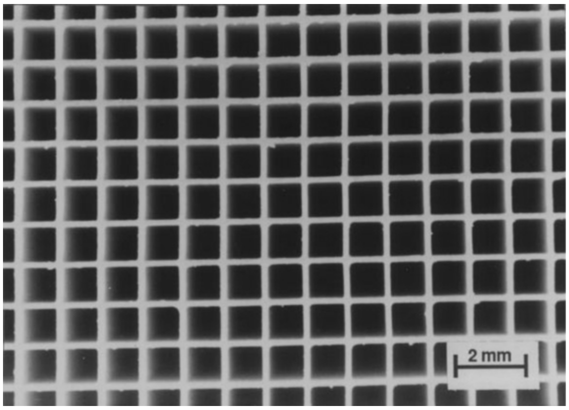


Figure 2
Structure of a ceramic honeycomb material.

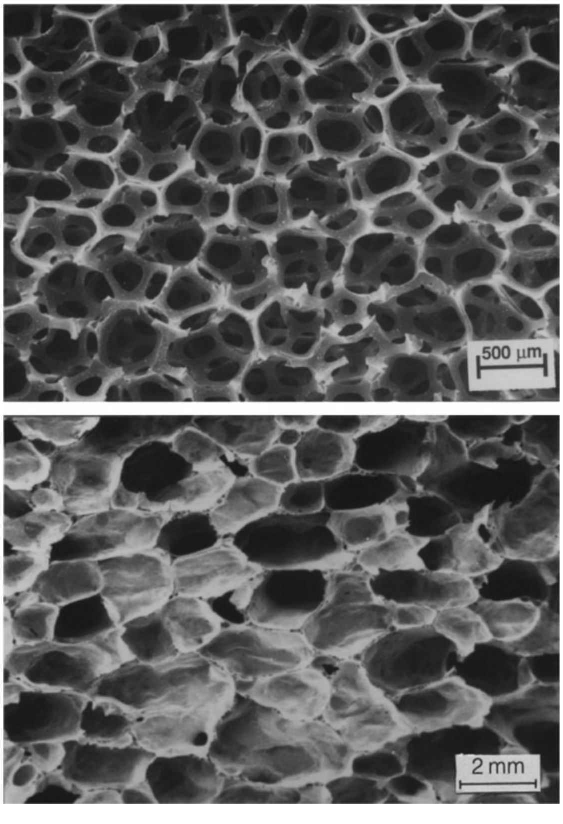


Figure 3
Comparison of the open-cell (top) and closed-cell structures in ceramic foams.

Cellular structures abound in nature and although they can be complex, they are often aesthetically pleasing and intriguing. It is reasonable to deduce that these natural cellular structures are not a result of random events but rather a careful evolutionary optimization process. These structures must fulfill a variety of functions but clearly one of these is the ability to withstand various types of mechanical forces. Indeed, for some stress geometry types, porous materials can provide maximum stiffness or strength for a given weight. The idea of producing (biomimetic) materials that imitate nature has been an important thrust in materials research.

2. *Properties of High-porosity Ceramics*

An important property of a high-porosity ceramic is its relative density, i.e., the density of the bulk material (ρ) normalized by the theoretical density of solid (ρ_s). Here it is worth emphasizing another important point with respect to the structure. The structure should be thought of at two levels, the first will be termed the macrostructure and refers to the arrangement of the fibers or cells. The finer structure that should be considered is the structure within the solid material and this will be termed the microstructure. The complex nature of the macrostructure usually leads workers to define a unit cell and its possible deformation modes are then analyzed. For example, for open-cell ceramic foams in which cell edge bending is predominant, one obtains

$$E = C_1 E_s \left(\frac{\rho}{\rho_s}\right)^2 \tag{1}$$

$$\mu = C_2 \mu_s \left(\frac{\rho}{\rho_s}\right)^2 \tag{2}$$

$$\sigma_{fc} = C_3 \sigma_{fs} \left(\frac{\rho}{\rho_s}\right)^{3/2} \tag{3}$$

$$T = C_4 \sigma_{fs} \sqrt{L} \left(\frac{\rho}{\rho_s}\right)^{3/2} \tag{4}$$

where E is Young's modulus, μ is the shear modulus, σ_{fc} is the compressive strength, T is the fracture toughness of the bulk foam, σ_{fs} is the mean flexural strength of the cell struts, L is the cell size, C_i are geometric constants, and the subscript "s" refers to the strut properties. One needs to be careful in using these equations as the strut properties may vary with cell size and density. For linear elastic materials, Eqn. (4) allows the tensile strength, σ_{ft}, to be written as

$$\sigma_{ft} = C_5 \sigma_{fs} \left(\frac{L}{a}\right) \left(\frac{\rho}{\rho_s}\right)^{3/2} \tag{5}$$

where a is the critical crack size. For this type of analysis for tensile strength, cracks are considered in terms of the number of neighboring broken struts. Thus, the lower limit of crack size is equivalent to the cell size. Under these circumstances, the tensile and compressive strengths are equivalent. Expressions similar to Eqns. (1)–(5) have been developed for closed-cell ceramics (see Gibson and Ashby 1997) and for fibrous ceramics (see Green 1984). For the latter case, the microstructural scaling parameter was fiber radius rather then cell size and it was necessary to account for the elastic and toughness anisotropy. Anisotropy can also occur in cellular ceramics but it is not usually as pronounced as in the fibrous ceramics. It is interesting to note that cell size (or fiber size) is not a significant factor in determining the mechanical properties. Even in Eqns. (4) and (5), strut strengths often decrease with increasing cell size as a result of the failure statistics involved in "finding" flaws in a stressed volume. The only key parameters that seem to be determined by cell size are permeability and high-temperature thermal conductivity. Thus, it is found, for example, that thermal conductivity and hence thermal shock resistance can decrease significantly with decreasing cell size.

Bibliography

Brezny R, Green D J 1994 Mechanical behavior of cellular ceramics. In: *Materials Science and Technology—A Comprehensive Treatment*. VCH, Weinheim, Germany Chap. 9

Gibson L J, Ashby M F 1997 *Cellular Solids: Structure and Properties*. Cambridge University Press, Cambridge

Green D J 1984 Mechanical behavior of Space Shuttle thermal protection tiles. In: Murr L E (ed.) *Industrial Materials Science and Engineering*. Dekker, New York, pp. 123–43

Vedula V R, Green D J, Hellmann J R 1999 Thermal shock resistance of ceramic foams. *J. Am. Ceram. Soc.* **82** (4), 649–56

D. J. Green
Pennsylvania State University, University Park, Pennsylvania, USA

Portland Cements

Cement is the binder used to make concrete mortars, stuccos and grouts for all types of building and construction. Although cement is a generic term and can be applied to many inorganic and organic materials, by far the most widely used and most versatile cement is Portland cement. This article is primarily concerned with Portland cement, but a brief description of other inorganic cements is also given. Unlike gypsum plaster, Portland cement is a hydraulic cement; i.e., it sets and hardens under water and hence can be used safely in all structures in contact with water. This article describes the manufacture and chemical composition of Portland cements used by the construction industry. It also summarizes the chemistry and microstructure of hydrated cement paste which is the binder in concrete and masonary construction.

The ancestry of Portland cement can be traced back to Greek and Roman times, but the year 1824 is generally considered to mark the origin of modern Portland cements when the name "Portland" was used in a patent by Joseph Aspdin to emphasize the similarity of concrete made with his cement to Portland limestone, which was a favored building material at the time. By the 1860s the Portland-cement industry was flourishing in the UK and Europe, but the first patent in the US was not taken out until 1870 by David Saylor, Today about 90 Mt are produced annually in the US.

1. Manufacture

Portland cement is commonly made from limestone and clay, two of the most widespread and abundant materials. The principal constituents are thus lime (CaO) and silica (SiO_2) which, when subjected to high temperatures ($\sim 1450\,^{\circ}C$), form the hydraulic calcium silicates that are the active compounds in cement. Other calcareous materials (such as muds, marls, and shell deposits) or argillaceous materials (such as slates, shales, and coal ashes) are used in some cases.

The overall manufacturing process is shown schematically in Fig. 1. The various steps are: (i) quarrying and crushing the raw materials, (ii) proportioning and fine grinding of the crushed materials, (iii) burning in a rotary kiln at high temperature, and (iv) grinding the resulting clinker to a fine powder with the addition of a small amount of gypsum. Although wet milling and proportioning with the aid of slurries (the *wet process*) is traditional and is found

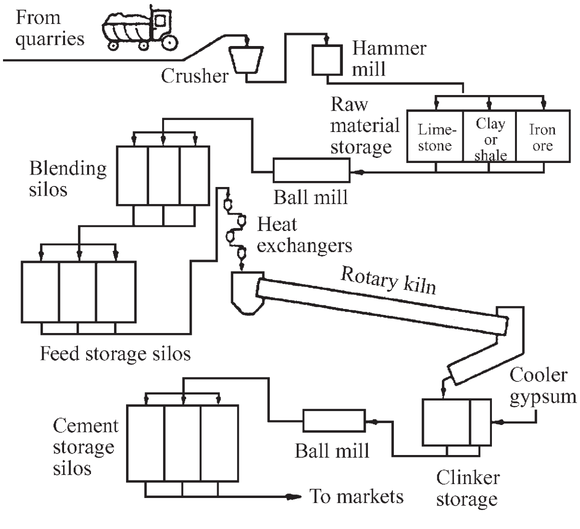


Figure 1
Schematic layout of a modern dry-process cement plant.

in older plants in the US, modern plants all use dry grinding and proportioning (the *dry process*). Although it was made possible by improvements in grinding mills, in particle-size classification, and in dust collection, the adoption of the dry process has been motivated largely by energy efficiency.

In the wet process, large amounts of energy are required to evaporate the water, and the long kilns, which must be used to attain large throughputs of materials (Fig. 2), have high heat losses. In addition, the dry process lends itself to the use of more efficient heat exchangers, such as the suspension preheater, in which the incoming kiln feed draws heat from the exit gases in a series of cyclones. As a result, dry-process rotary kilns are much shorter than those used in the wet process (Fig. 2), heat losses are much reduced, and throughputs are higher. Further savings can be realized by precalcining the hot feed in a flash furnace (precalciner) before it enters the kiln.

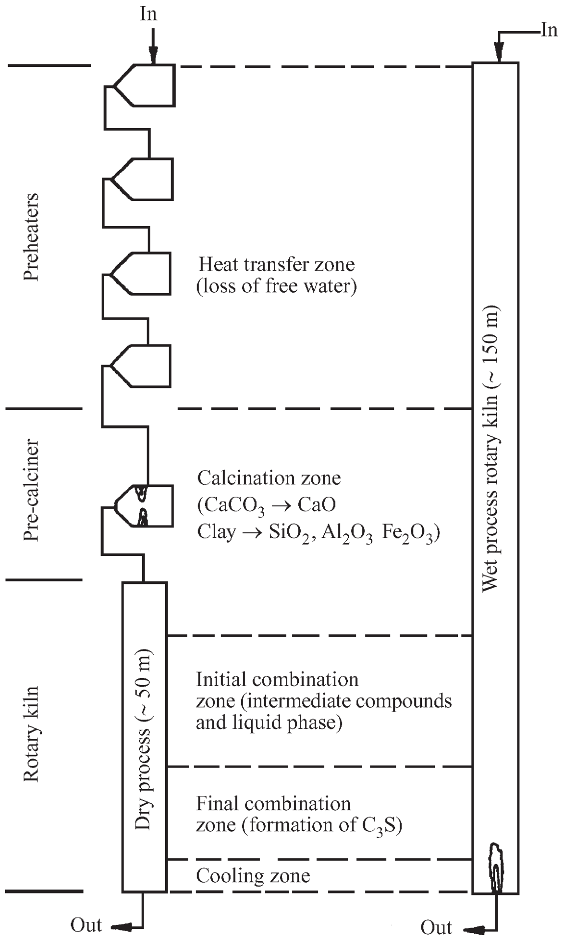


Figure 2
Comparison of kiln reactions in the wet and dry processes.

Typical energy consumptions in cement manufacture are given in Table 1. Average overall energy consumptions ($\mathrm{MJ\,kg^{-1}}$) for countries with cement industries employing various processes are as follows; Japan, 4.5; FRG, 4.2; and US, 5.7. Theoretically a value of $1.7\,\mathrm{MJ\,kg^{-1}}$ is calculated from the basic chemical processes involved; in practice, modern plants now have energy consumptions approaching $3.5\,\mathrm{MJ\,kg^{-1}}$. The average figure for the US is relatively high because of a preponderance of older plants built when energy was inexpensive.

As calcination of the raw materials progresses, chemical combination takes place in the hotter zones of the rotary kiln (see Fig. 2). Formation of intermediate compounds by sintering occurs first, until at 1335 °C a quarternary liquid phase appears with a composition close to that predicted from the C-S-A-F phase diagram. ($C = CaO$, $S = SiO_2$, $A = Al_2O_3$, and $F = Fe_2O_2$ in conventional cement chemists' notation; other symbols employed include $S = SO_2$, $H = H_2O$, and $K = K_2O$.) The amount of liquid formed depends on the amount of A and F present in the raw feed and on the kiln temperature.

Generally, about 25% of the total mix is liquid and hence the process is neither a complete fusion or a pure solid-state sintering; it is called *clinkering*. The liquid phase allows chemical combination to go to completion in a relatively short time. Much C_2S has formed by solid-state reactions before the liquid appears but its formation is completed in the liquid. The final reaction between C_2S and lime to form C_2S takes place wholly within the liquid at 1450–1500 °C. The rate of combination depends on the rate of dissolution of lime in the clinker liquid, which is dependent on both particle size and temperature.

On leaving the hottest zone, the clinker cools to about 1100–1200 °C. Then it is air-quenched as it exits the kiln. However, cooling is usually too rapid to allow true equilibrium to be maintained. The rate of cooling also controls the crystal growth of the calcium silicates and the crystallization of liquid phase to form C_3A, C_4AF, and additional C_2S ($C_3A = 3CaO.Al_2O_3$, etc.), slow cooling results in the formation of larger and less-reactive crystals. The structure of a clinker can be studied by optical microscopy and used to predict clinker performance. After cooling, the clinker may be stored for future grinding or it may be ground immediately. Finish grinding takes place in ball mills equipped with particle-size

Table 1
Average energy consumption in cement manufacture ($\mathrm{MJ\,kg^{-1}}$).

Kiln operation	Wet process	Dry process
Fuel consumption	6.5	4.7
Electrical energy	0.5	0.6

classifiers that avoid over-grinding. At this time, small quantities of gypsum are added to control the early hydration reactions of C_3A and thus optimize cement performance.

2. *Composition*

Portland cement consists of four major clinker compounds, C_3S, C_2S, C_3A, and C_4AF together with the gypsum added during grinding. The relative portions of these compounds can be changed to optimize specific properties of the cement (discussed in more detail below). Table 2 shows typical compound compositions, which are calculated assuming (incorrectly) ideal stoichiometries and chemical equilibrium in the kiln (the Bogue calculation).

All the clinker compounds form solid solutions with small amounts of every element that is in a clinker. In addition to the four major oxides, MgO, K_2O, and Na_2O are ubiquitous impurities in most raw materials. Impure C_3S is called *alite*: it reacts more rapidly with water than does pure C_2S. C_2S exists as four major polymorphs: γ, β, α', and α. At ambient temperatures, the nonhydraulic γ polymorph is thermodynamically stable but, in cement, impurities stabilize the hydraulic β polymorph. Impure C_3A contains considerable quantities ($\sim$10 mol.%) of iron, which may make it less reactive than the pure compound. High levels of sodium substitution modify its crystal structure from orthorhombic to cubic with a corresponding decrease in reactivity.

C_4AF does not have an exact stoichiometry; continuous solid solution between A and F allows its composition to vary from about C_6A_2F to C_6AF_2. For this reason C_4AF is usually called the ferrite phase. Although in many cements the composition of the ferrite phase is close to C_4AF, in those that have very low C_3A contents, the composition is closer to C_6AF_2. As the iron content increases, the reactivity of the ferrite phase decreases. It is the ferrite phase that gives cement its color: when MgO is present in the ferrite phase, it becomes black and the cement gray.

If the MgO content is low, or if the ferrite phase contains appreciable ferrous iron (due to a reducing atmosphere in the kiln), the cement is browner. In addition, crystalline MgO (periclase) is present in most clinkers, typically in the range 2–4 wt.%. There are limits placed on the amount of MgO, because its slow hydration can cause disruptive expansions in the hardened paste. In clinkers that are produced using high sulfur coal as a fuel, alkali sulfates, such as $K\bar{S}$ or $C_2K\bar{S}_3$ sometimes occur; these compounds may affect setting behavior and can lower the 28-day strength.

Cements can be classified according to the ASTM types given in Table 2. These classifications are based mostly on performance, not composition, although in general cements of the same type have broadly similar compositions. Over 90% of all Portland cement made in the USA is type I. When higher strength is required at early ages (1–3 days) a type III can be used. This is more finely ground so that its overall hydration is faster. Type IV cement can be used to avoid thermal cracking in mass concrete, by lowering the heat of hydration and hence the internal temperature of the mass but are no longer manufactured in the US. In the presence of high concentrations of sulfates in ground waters or soils, the hydration products of C_3A may undergo further reactions that result in disruptive expansion. The resistance of concrete to sulfate attack can be improved by lowering the C_3A content in a type V cement. Type II cement has properties intermediate between those of type I and type V.

Aside from the five ASTM types of Portland cement, other special cements are available. Blended cements, containing reactive siliceous materials (pozzolans), or hydraulic slags mixed with type I, are used in most countries as alternatives to types IV and V. On hydration, they form more C–S–H and less calcium hydroxide. Expansive cements are modified Portland cements that counteract the effects of drying shrinkage by forming large amounts of ettringite to generate restrained expansion. Some special cements use ettringite to provide high early strength. Masonry cements and oil-well cements are Portland cements, modified to optimize the special requirements of these applications. Calcium aluminates (high-alumina cements) are used principally for refractory concretes. They cannot be used for structural purposes because of potential strength loss

Table 2
Average compound composition of Portland cement (wt.%).

ASTM type	Description	C_3S	C_2S	C_3A	C_4AF	$C\bar{S}_2$
I	General purpose	55	17	10	7	6
II	Moderate sulfate resistant	55	20	6	10	5
III	High early strength*	55	17	9	8	7
IV	Low heat of hydration	35	40	4	12	4
V	Sulfate resistant	55	20	4	12	4

Note: *This cement is usually ground more finely: $\sim 550\,m^2\,kg^{-1}$ compared to $380\,m^2\,kg^{-1}$

Table 3
Hydration of clinker compounds.

Compound	Hydration reactions	Heat of reaction (kJ mol^{-1})	Heat of reaction J g^{-1}
Alit	$2C_3S + 11H \rightarrow C_3S_2H_8 + 3CH$	−114	−500
Belite	$2C_2S + 9H \rightarrow C_3S_2H_8 + CH$	−43	−250
C_3A	$C_3A + 3C\bar{S}H_2 + 26H \rightarrow C_6A\bar{S}_3H_{32}$ $2C_3A + C_6A\bar{S}_3H_{32} + 4H \rightarrow 3C_4A\bar{S}H_{12}$	−365	−1350
Ferrite	$C_4AF + 15H \rightarrow C_4(A_3F)\bar{S}H_{12} + (A,F)H_3$	−403	−200

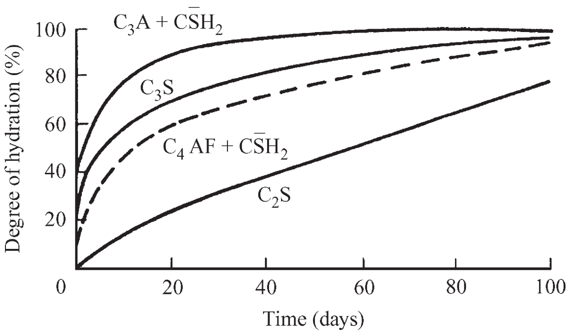


Figure 3
Rate of hydration of the clinker compounds.

during service. Magnesium oxychloride (Sorel) cements can be used only for interior applications because they do not resist moisture.

3. Hydration

The setting and hardening of concrete is the result of complex chemical reactions between cement and water (hydration) in the cement paste binder. The overall reactions are summarized in Table 3. Both alite and belite form the same hydration products: an amorphous calcium silicate hydrate ($C_3S_2H_8$) and crystalline calcium hydroxide (CH). (The designated stoichiometry of the calcium silicate hydrate is only an approximation, since its composition is variable. It is usually referred to as C–S–H, which implies no definite stoichiometry or regular crystal structure.) The rate of hydration (Fig. 3) and the heat of hydration of the two silicates are quite different. The silicates are the components responsible for the strength of cement pastes, but in the first seven days the alite content controls the development of strength. The hydration of alite has an initial induction period that keeps concrete fluid for 1–2 hours and allows it to be handled and placed conveniently.

In the presence of gypsum, C_3A and the ferrite phase form calcium sulfoaluminate hydrates. The first product is the trisulfate form, ettringite, but when all gypsum has been consumed ettringite converts to the monosulfate form. Although C_3A and the ferrite phase contribute little to strength, they are still important components. When $C_4A\bar{S}H_{12}$ is exposed to an external source of sulfate ions, it will convert back to ettringite with concomitant expansion.

Interactions occur between the clinker compounds when they are hydrating in cement. For example, belite hydrates faster in the presence of alite, C_3A and the ferrite phase compete for the gypsum during their hydration, and gypsum accelerates the hydration of alite. During its formation, C–S–H can incorporate considerable quantities of alumina, sulfate, and Fe_2O_3 into solid solution. Thus less sulfoaluminates are formed during hydration than theory predicts; perhaps as little as half the expected amount.

It is important to know the effect of temperature on the hydration process because often concretes are cured at elevated temperatures. During the first 24 hours, the hydration of cement is very sensitive to temperature; apparent activation energies of 40 kJ mol^{-1} have been reported for cement or alite. Thus, at low temperatures the rate of hydration, and hence of strength development, is very much reduced, while at high temperatures the reaction is accelerated. The calcium sulfoaluminate hydrates become unstable at 70–90 °C, while above 100 °C, C–S–H will convert to crystalline compounds. In the temperature range of autoclaved concrete products (150–200 °C under pressurized steam), C–S–H rapidly transforms to α-C_2SH. This results in low strength because the transformation increases porosity, but when finely divided silica is added, C–S–H converts to 1.1 nm tobermorite ($C_5S_6H_5$), which has good cementing properties.

4. Structure of Hardened Cement Paste

The transition of a hydrating cement paste from a fluid to a rigid material is a continuous process resulting from progressive hydration of the calcium silicates. Ettringite may control rheological properties, but does not form a continuous matrix because C_3A and the ferrite phase are minor components. Ettringite crystallizes mostly at the surface of the

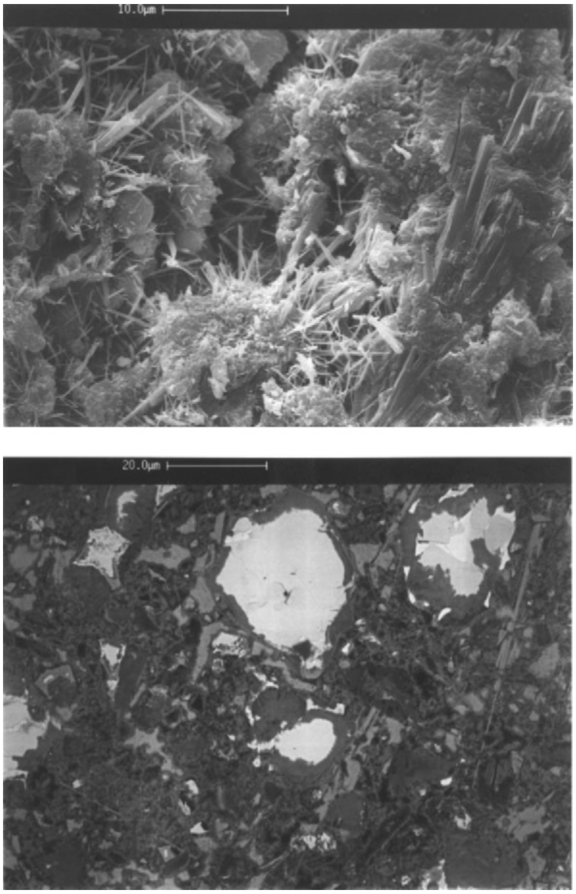


Figure 4
Microstructure of hardened cement paste revealed by scanning electron micrographs. (Courtesy of Dr. Paul Stutzman, National Institute of Standards and Technology).Upper panel: fracture surface of a 7-day old paste. Ettringite needles are growing into waterfilled pores and monolithic $Ca(OH)_2$ on the right-hand side. (× 2500)Lower panel: polished epoxy-impregnated 28-day old cement paste. Bright particles of unreacted cement are surrounded by dark grey C–S–H. The matrix between the particles is an intimate mixture of C–S–H (dark grey) and $Ca(OH)_2$ (light grey) (× 1000).

cement grains but also grows in the pore space (Fig. 4). Under special conditions crystallization of larger amounts of ettringite can cause premature stiffening because of physical interlocking of the needle-like crystals. The continuous matrix of hardened paste is formed as C–S–H coatings at the surface of calcium silicate particles merge as hydration proceeds. Calcium hydroxide grows in the water-filled voids between the hydrating particles. The approximate composition of the paste is given in Table 4.

C–S–H is amorphous colloidal material with high surface area and small particle size; these features can be seen only with an electron microscope. It has a large volume of intrinsic microporosity (~ 25 vol.%), which lowers its intrinsic strength and causes volume instability on drying. To a large extent, the material properties of concrete reflect the properties of C–S–H. In contrast, massive crystals of calcium hydroxide can be readily seen, even by optical microscopy. Calcium hydroxide contributes to strength because it fills voids, but it is highly soluble. The presence of soluble alkali salts makes the concrete pore

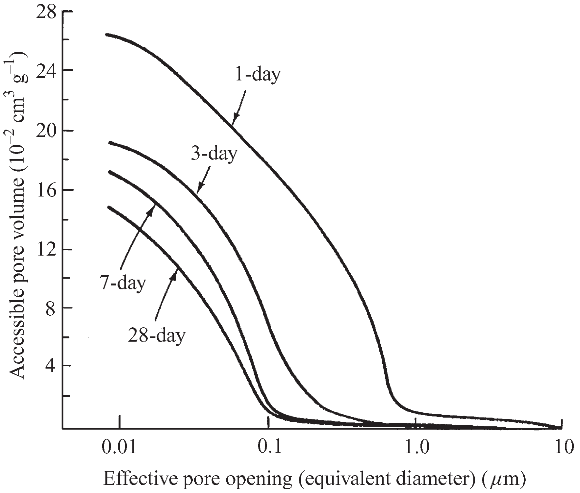


Figure 5
Accessibility of the pore volume of hydrated cement paste as measured by mercury intrusion porosimetry.

Table 4
Composition of hardened paste (water–cement weight ratio, w/c = 0.5).

Volume component	Percent	Remarks
C–S–H	55	Continuous matrix: high intrinsic porosity; amorphous
CH	20	Large crystals (0.01–1 mm)
Calcium sulfoaluminate hydrates	10	Small crystals (1–10 µm)
Capillary porosity	~ 15	Depends on the amount of water used

Table 5
Influence of paste porosity on concrete properties.

Pore component	Size	Effect on properties
Capillary pores		
Macropores	$>0.1\,\mu\text{m}$ (100 nm)	Permeability coefficient; Ionic diffusivities Strength (also total porosity)
Large mesopores	10–100 nm	Permeability coefficient; Ionic diffusivities Drying shrinkage
Gel pores		
Small mesopores	2.5–10 nm	Drying shrinkage; Creep
Micropores	<2.5 nm	Drying shrinkage; Creep

Table 6
Mineral admixtures used in concrete.

Admixture	Form of reactive silica	Reaction
Fly ash– Class F	Aluminosilicate glass	Pozzolanic
Class F	Calcium aluminosilicate glass	Pozzolanic (cementitious)
Silica fume	Amorphous silica ($\sim 20\,\text{m}^2\text{g}^{-1}$)	Pozzolanic
Granulated blast furnace slag	Calcium aluminiosilicate glass	Cementitious + pozzolanic
Calcined clay	Amorphous silica (and alumina) ($\sim 10\,\text{m}^2\text{g}^{-1}$)	Pozzolanic
Rice husk ash	Amorphous silica ($\sim 60\,\text{m}^2\text{g}^{-1}$)	Pozzolanic

solution very alkaline, and this can reduce the durability of concrete in certain circumstances. The calcium sulfoaluminate hydrates play a relatively minor role in the behavior of hardened pastes, except when sulfates are present.

All properties of concrete are affected by the volume of capillary pores in the hardened cement paste. The pores are interconnected, allowing water and dissolved species to penetrate the concrete, with potential adverse consequences for service life (i.e., durability). Transport properties such as water permeability and ionic diffusivity are controlled by the size and volume of pores. A pore size distribution can be obtained using mercury intrusion porosimetry (Fig. 5). The volume of pore space accessible through a pore diameter $>0.1\,\mu\text{m}$, and the pore diameter corresponding to the maximum rate of intrusion are the main parameters controlling transport. The capillary pore volume is controlled by the water–cement ratio and the extent of hydration during curing. Capillary pores also affect other properties as well (see Table 5).

4.1 Modification of Microstructure

The microstructure can also be modified by the use of reactive aluminosilicate (mineral admixtures), such as fly ash, silica fume, or granulated blast furnace slag. These compounds contain amorphous silica and react chemically to convert calcium hydroxide to C–S–H (see Table 6). This the so-called *pozzolanic reaction*. In addition, slag will react cementitiously to form C–S–H directly. The C–S–H formed by mineral admixtures has a lower C/S molar ratio than that formed by direct hydration.

Increasing the volume factor of C–S–H in the microstructure reduces the overall porosity because C–S–H has a lower specific gravity than $Ca(OH)_2$. The proportion of micropores and small mesopores also increases. As a result such pastes have enhanced resistance to the transport of water and dissolved species, thereby creating more durable concrete. The reduction of the volume fraction of leachable $Ca(OH)_2$ and of the pore solution also enhances durability.

5. *Concluding Remarks*

Portland cement is a calcium silicate cement formed by high temperature reactions between limestone and clay. The reaction of calcium silicates with water (hydration) are responsible for the strength and durability of concrete. Small quantities of alumina and iron oxide which act as fluxing agents during manufacture form calcium aluminates, which also react with water. The hydration products are arranged physically to form a solid structure with a network of small pores. Hydration chemistry and the resulting microstructure can be modified by the addition of reactive aluminosilicates.

Bibliography

Barnes P (ed.) 1984 *Structure and Performance of Cement*. Applied Science, Barking, UK
Bye G C 1983 *Portland Cement: Composition, Production and Properties*. Pergamon, Oxford
Francis A J 1977 *The Cement Industry 1796–1914: A History*. David and Charles, Newton Abbot, UK
Ghosh S N (ed.) 1983 *Advances in Cement Technology*. Pergamon, Oxford
Kurdowski W 1991 In: Ghosh S N (ed.) Cement and Concrete Science and Technology. Vol. 1, Part 1, ABI Books, New Delhi, India, pp. 58–98 and pp. 201–36
Lea F M, Hewlett P C 1998 *Lea's Chemistry of Cement and Concrete*, 4th edn. Arnold, London
Mehta P K, Montiero P J M 1993 *Concrete: Microstructure, Properties, and Materials*. McGraw-Hill, New York
Mindess S, Young J F 1981 *Concrete*. Prentice Hall, Englewood Cliff, New Jersey, USA
Skalny J P (ed.) 1989 *Materials Science of Concrete. Vol. I*. American Ceramic Society, Westerville, OH, USA
Taylor H F W 1997 *Cement Chemistry*, 2nd edn. Thomas Telford, London

J. F. Young
University of Illinois, Urbana, Illinois, USA

Precipitation from Austenite

In pure iron its f.c.c phase, austenite or γ phase, is stable between 912 °C and 1394 °C. This situation is significantly changed by the addition of alloying elements. For instance, the addition of C expands the austenitic phase field, allowing a relatively wide range of temperatures and C contents within which the austenite remains stable. In iron–carbon alloys and also in low alloy steels the austenitic field is mainly important for the steel processing. For example, "austenitizing" is normally an important step in heat-treating. In addition, hot working is normally carried out in the austenitic field. The addition of a large concentration of alloying elements, notably γ phase stabilizers such as C, Ni, and Mn, can cause the steel to remain at least partly austenitic down to the room temperature.

The microstructure of the austenitic steels at room temperature following rapid cooling from the solution annealing temperature consists primarily of a polycrystalline matrix with low stacking fault energy. However, even after rapid cooling, austenite is rarely the only phase present. Nonmetallic inclusions and possibly some undissolved carbides or some carbides and nitrides, "primary precipitates," which precipitate from the melt during solidification can also be present. Given sufficient time and temperature, other phases often precipitate from the austenite, "secondary precipitates," such as carbides, nitrides, borides, and also intermetallic phases. Of course, ferrite and martensite formation can sometimes occur, but these are the subjects of other articles in this encyclopedia (see *Ferrous Alloys: Overview, Martensite*). In this article, the most common precipitation reactions from the austenite are discussed, emphasizing precipitation from high alloy iron-base austenites. High-temperature precipitation of carbides/nitrides resulting from microalloying additions to low alloy austenite is also presented briefly.

1. Precipitation from High Alloy Austenite

Some common high alloy austenitic systems are: Fe–Cr (15–28 wt.%)–Ni (6–35 wt.%) or Fe–Cr (15–28 wt.%)–Ni (8–35 wt.%)–Mo (1–7 wt.%) that form the basis of austenitic stainless steels; Fe–Cr (11–30 wt.%)–C (1.8–3.6 wt.%) that form the basis of the high-chromium white cast irons; and Fe–Mn 12–20 wt.%)–C (1–2 wt.%) with Mn:C ratio about 10 : 1, these high-manganese austenitic steels have been employed as wear-resistant materials since the original austenitic manganese steel (Fe–1.2 wt.%C–12%Mn) was invented by Sir Robert Hadfield in 1882. These alloys, in which austenite is the main phase at room temperature, were chosen as representative of the

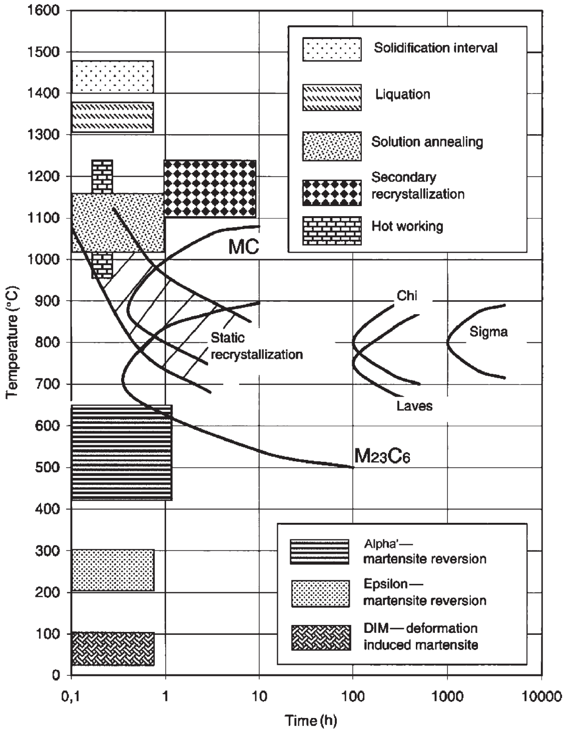


Figure 1
Main heat treatments and transformations that occur in austenitic stainless steels between room temperature and the liquid state (after Padilha, Plaut, and Rios).

Table 1
The crystal structure, lattice parameters, composition, and occurrence of carbides, nitrides, borides, and intermetallic phases in austenitic alloys.

Phase	Crystal structure	Typical lattice parameters (nm)	Composition	Example of occurrence
Carbides				
MC	f.c.c.	$a_0 = 0.4328$	Ti(C,N); Nb(C,N)	AISI 321; 347
M_6C (η)	f.c.c.	$a_0 = 1.095$–1.128	$(Fe,Cr,Mo)_6C$	AISI 316
$M_{23}C_6$	f.c.c.	$a_0 = 1.057$–1.068	$(Fe,Cr,Mo)_{23}C_6$	AISI 316; 304
M_7C_3	Pseudo-hexagonal	$a_0 = 1.398$ $c_0 = 0.4541$–0.4511	$(Fe,Cr)_7C_3$	Fe–18 wt.%Cr–3 wt.%C Fe–12 wt.%C–1.1 wt.%C
M_3C	Orthorhombic	$a_0 = 0.452$ $b_0 = 0.509$ $c_0 = 0.674$	$(Fe,Mn,Cr)_3C$	
Nitrides				
MN	f.c.c.	$a_0 = 0.4240$	TiN	AISI 321
M_2N	Hexagonal	$a_0 = 0.478$ $c_0 = 0.444$	$(Cr,Fe)_2N$	Fe–25 wt.%Cr–5.5 wt.%Ni–0.87 wt.%N
Borides				
M_2B	Orthorhombic	$a_0 = 1.458$ $b_0 = 0.738$ $c_0 = 0.4245$	$(Cr,Fe)_2B$	AISI 304 + B
M_3B_2	Orthorhombic		$(Cr,Fe,Mo)_3B_2$	AISI 316 + B
Intermetallic phases				
σ	b.c. tetragonal	$a_0 = 0.87$–0.92 $c_0 = 0.4544$–0.48	$(Fe,Ni)_x\,(Cr,Mo)_y$	AISI 304; 316
χ	b.c.c.	$a_0 = 0.8807$–0.892	$(Fe,Ni)_{36}Cr_{18}Mo_4$	AISI 316
Laves (η)	Hexagonal	$a_0 = 0.473$–0.474 $c_0 = 0.727$–0.785	Fe_2Mo; Fe_2Nb; Fe_2Ti	AISI 316
G	f.c.c.	$a_0 = 1.122$–1.124	$Ni_{16}Nb_6Si_7$;$Ni_{16}Ti_6Si_7$	AISI 316
R	Hexagonal	$a_0 = 1.093$ $c_0 = 1.9342$	(Fe,Cr,Mo)	22 wt.%Cr–8 wt.%Ni–3 wt.%Mo
γ'	f.c.c.	$a_0 = 0.3605$	$Ni_3(Al,Ti)$	Alloy 800

(Weiss and Stickler 1972, Peckner and Bernstein 1977, Harries 1981, Marshall 1984, Davis 1994, Villars *et al.* 1995).

diverse compositions that high alloy austenite might have. Austenite with even higher alloy contents can be found in the "superaustenitic" austenitic stainless steels containing high amounts of Cr, Ni, Mo, and N, resulting in an iron content close to or less than 50 wt.% (Heino *et al.* 1999, Lee *et al.* 2000). "Superaustenitic" austenites will not be treated here but the main difference between these alloys and those treated here, from the viewpoint of precipitation, is a greater tendency of these alloys to form intermetallic phases owing to their higher alloy content. Precipitation from high alloy austenite has been the subject of two recent reviews (Sourmail 2001, Padilha and Rios 2002), another recent review discusses the effect of deformation on the precipitation (Padilha *et al.* 2003). The main transformations that occur in a high alloy austenite are summarized in Fig. 1, of which only the precipitation reactions are discussed in this article.

The most common carbides/nitrides/borides and intermetallic phases that precipitate from high alloy austenite are listed in Table 1. The corresponding crystal structure, composition range, and typical lattice parameter are also shown, together with examples of common systems of engineering relevance in which such a phase is likely to form.

1.1 Carbide Precipitation

$M_{23}C_6$ ($M = Cr, Fe, Mo$)

The C solubility in Fe–Cr–Ni austenite is 0.15 wt.% at 1100 °C but less than 10 ppm at 600 °C in austenitic stainless steels. Therefore, when the steel is aged at 550–900 °C, precipitation of $M_{23}C_6$ carbides is unavoidable (Weiss and Stickler 1972, Peckner and Bernstein 1977). The precipitation of $M_{23}C_6$ occurs mainly on grain boundaries, incoherent twin boundaries, coherent twin boundaries, and dislocations. The $M_{23}C_6$ carbide precipitates normally exhibits the following orientation relationship with the matrix:

$$\{100\}\gamma \| \{100\}M_{23}C_6 \quad \text{and} \quad \langle 100\rangle\gamma \| \langle 100\rangle M_{23}C_6$$

Extensive grain boundary precipitation of $M_{23}C_6$ is undesirable, since the regions in the neighborhood of the grain boundaries become impoverished in Cr and therefore susceptible to intergranular corrosion and to the formation of martensite upon cooling. This typical situation is illustrated in Fig. 2. The main engineering consequences of $M_{23}C_6$ precipitation are degradation of intergranular corrosion resistance and reduction in tensile properties, especially ductility and toughness (Marshall 1984). In order to prevent such precipitation, one can either lower the C content of the alloy or add elements that are able to form the more stable MC type carbides such as Nb or Ti. Although these measures can reduce the amount of $M_{23}C_6$ precipitation and prevent sensitization, it is rarely possible to avoid completely $M_{23}C_6$ precipitation on account of its fast precipitation kinetics.

(a) M_6C (M = Mo, W, Fe, Nb, V)

The carbides of the type M_6C are often found in austenitic stainless steels containing Mo, W, and Nb, especially Mo. The M_6C carbide invariably contains more than one metallic element and is usually represented by the formulas A_3B_3C or A_4B_2C. Nitrogen additions are believed to favor the M_6C precipitation to the detriment of $M_{23}C_6$, because the former is able to dissolve more N than the latter. As $M_{23}C_6$ absorbs Mo during aging and the Mo content in $M_{23}C_6$ exceeds a certain threshold, an *in situ* $M_{23}C_6 \rightarrow M_6C$ transformation is possible (Weiss and Stickler 1972, Peckner and Bernstein 1977). It is important to mention that comparatively less attention has been given to this type of carbide in the literature, because it is not found (or sometimes found only in small quantities) in most of the stainless steels. With the advent of the "superaustenitic" stainless steel, which contains high levels of Mo and N, it is likely that more investigations will be done on this type of carbide and its effects on properties.

MC (M = Ti, Zr, Hf, V, Nb, Ta)

The MC carbides are very stable and may be present in the microstructure of the stabilized austenitic stainless steels, such as the Ti-stabilized AISI 321 and the Nb-stabilized AISI 347. In these steels, the formation of these carbides is aimed at reducing or preventing the precipitation of $M_{23}C_6$ as mentioned above. They basically show two types of distribution: (i) a coarse dispersion, 1–10 μm in size, of primary particles formed during solidification; and (ii) a fine dispersion, 5–50 nm in size, of secondary precipitates. In the stabilized steels, part of the primary carbides can be dissolved during the solution annealing heat treatment, at temperatures ranging from 1050 °C to 1150 °C, and them reprecipitate as fine secondary precipitates during the aging heat treatment or when these materials are subject to high-temperature applications. The MC carbide precipitation is predominantly intragranular, on dislocations and stacking faults. Figure 3 shows

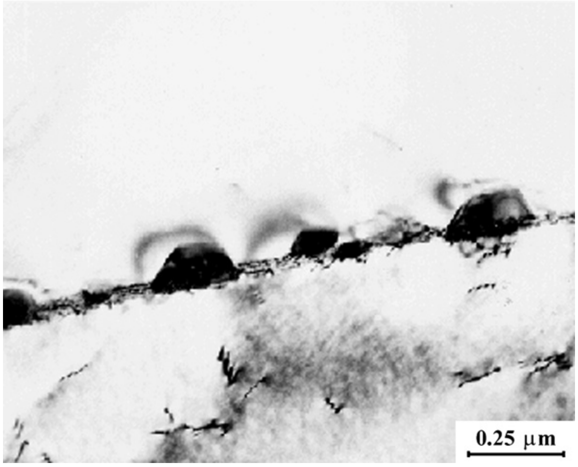


Figure 2
Grain boundary $M_{23}C_6$ precipitates in an austenitic stainless steel observed using transmission electron microscopy (after Padilha and Rios).

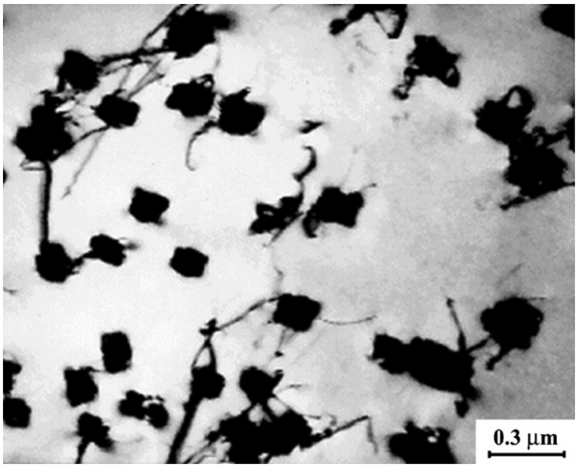


Figure 3
(Ti, Mo)C carbides precipitation on dislocations in a Ti-stabilized austenitic stainless steel observed by transmission electron microscopy (after Padilha and Rios).

(Ti,Mo)C precipitates on dislocations in a Ti-stabilized austenitic stainless steel. Although the lattice parameter differences between the matrix and the MC carbides are higher than 10%, a cube-on-cube orientation relationship is frequently found between the MC and parent austenite. This type of carbide has a particularly large resistance to coarsening (Harries 1981).

(b) M_3C (M = Fe, Mn, Cr) and M_7C_3 (M = Cr, Fe)

The carbides of the type M_3C are not found in austenitic stainless steels and carbides of the type M_7C_3 (M = Cr, Fe) can be found only under particular conditions, for very high C:Cr ratios, for example, during carburizing. Both carbides can, however, be found in high-Cr white cast irons that are traditionally employed in applications where there is a need for a high resistance to abrasive wear. They contain 11–30 wt.% of Cr and 1.8–3.6 wt.% of C, as well as additions of Mo, Mn, Ni, and Cu.

The as-cast microstructure consist of a dispersion of primary and/or eutectic M_7C_3 and sometimes M_3C carbides in an austenitic matrix (Tabrett *et al.* 1996). The as-cast microstructure can be significantly changed by austenite destabilization heat treatments (Tabrett *et al.* 1996, Powell and Bee 1996). These heat treatments are carried out at ~1000 °C for holding times of several hours. During destabilization treatments, there is precipitation of secondary carbides of types M_3C or M_7C_3 or even $M_{23}C_6$, decreasing the alloying element concentration in the matrix, especially C. The type of the secondary carbide formed depends on the matrix composition and the destabilization temperature. In alloys containing a high Cr content (>25 wt.%), there is a tendency to precipitate $M_{23}C_6$ carbides, while in alloys with a Cr content of 15–20 wt.% there is a predominance of M_7C_3 carbides. These secondary carbides predominantly precipitate on slip bands and subgrain boundaries within the austenite. The secondary carbide precipitation depletes the matrix of alloying elements and increases the M_s temperature. As a consequence, after the destabilization treatment, the matrix is mainly martensitic but it can contain up to 35 vol.% of retained austenite.

Most of the high-Mn austenitic steels have the manganese content in the range of 12–20 wt.% and a C : Mn ratio of about 1 : 10. The solution heat treatment, performed in the range 950–1100 °C and followed by quenching, gives a microstructure free of carbides. When the steel is held at 350–850 °C, there is precipitation of M_3C carbides (Yang and Choo 1994). In the lower-temperature range, 350–500 °C, carbide precipitation occurs at grain boundaries and at crystalline defects within the grains. Above 500° C, pearlite formation can take place, depending on the Mn content.

1.2 Nitride Precipitation: MN (M = Zr, Ti, Nb) and $(Cr,Fe)_2N$

Nitrogen is added to stainless steels because it improves mechanical properties and corrosion resistance, and also because it is a strong austenite stabilizer. Up to 0.8 wt.%, N can be dissolved in austenitic stainless steels without causing cleavage cracking at cryogenic temperatures and loss of toughness.

The nitrides that form in the austenitic stainless steels can be grouped into two classes: (i) primary nitrides of type MN (M = Zr, Ti, Nb), formed in stabilized steels containing residual amounts of N (<0.1 wt.%) and (ii) secondary nitrides of the type M_2N (M = Cr, Fe), which precipitate in stainless steels containing high N (0.2–0.8 wt.%). The MN nitrides have the same crystalline structure as the MC carbides, but they are even more stable and do not dissolve during solution annealing. The solubility of N in Fe–Cr–Ni or Fe–Cr–Mn–Ni austenites decreases considerably with temperature. The $(Cr,Fe)_2N$ nitrides can precipitate continuously or discontinuously (Jargelius-Pettersson 1996, Machado and Padilha 1996). The N depletion of the matrix due to precipitation of nitrides can make the austenite unstable and it can make possible the formation of ferrite at aging temperatures (Machado and Padilha 1996).

Finally, it must be mentioned that N in solid solution delays the precipitation of phases which do not dissolve or dissolve very little N, such as $M_{23}C_6$, σ, χ, and Laves.

1.3 Boride Precipitation: $(Cr,Fe)_2B$ and $(Cr,Fe,Mo)_3B_2$

The positive influence of small additions, 10–80 ppm, of boron on the creep behavior and hot workability of austenitic stainless steels is well established. Nonetheless, if the boron content is too high, it may cause the formation of low melting point (1150–1225 °C) eutectics, which may cause hot shortness during hot working and welding (Marshall 1984). When held at 650–1050 °C, stainless steels containing more than ~30 ppm of B are susceptible to the precipitation of two types of orthorhombic borides: $(Cr,Fe)_2B$ and $(Cr,Fe,Mo)_3B_2$. M_2B is found more frequently and the M_3B_2 occurs only in steels with higher Mo contents. The $(Cr,Fe)_2B$ phase precipitates chiefly on the grain boundaries. When it precipitates within the grain, it is often found to form at the incoherent phase boundaries of primary carbides (Padilha and Schanz 1980). The beneficial effects of B are normally associated with B in solid solution. The volume fraction of borides is very small and has no significant effect on mechanical properties. Notwithstanding in nuclear applications the amount and the distribution of boron are important (Padilha and

Schanz 1980). The reason for this is that the ^{10}B isotope can react with neutrons producing He, which causes high-temperature embrittlement.

1.4 Precipitation of Intermetallic Phases

The precipitation of intermetallic phases from austenite is normally associated with undesirable consequences like matrix impoverishment in alloying elements such as Cr, Mo, and Nb and loss of ductility and toughness. Exceptions are the γ'-$Ni_3(Al,Ti)$ and, to a lesser extent, the Laves phases, Fe_2Nb, since their precipitation can result in precipitation hardening. Precipitation of intermetallic phases is slower than that of $M_{23}C_6$. As shown in Fig. 4, $M_{23}C_6$ can start to precipitate in less than an hour whereas intermetallic phases can take over a hundred hours to start.

(a) Sigma (σ) (Fe–Cr and Fe–Cr–Mo)

The σ-phase appears in several binary, tertiary, and quaternary systems such as Fe–Cr, Fe–Mo, Fe–V, Fe–Mn, Fe–Cr–Ni, Fe–Cr–Mo, Fe–Cr–Mn, and Fe–Cr–Ni–Mo. A typical σ phase chemical composition, found in the AISI 316 or 316L austenitic stainless steels, is the following (in wt.%): Fe = 55%, Cr = 29%, Mo = 11%, and Ni = 5% (Weiss and Stickler 1972). Alloying elements such as Cr, Mn, Mo, W, V, Si, Ti, Nb, and Ta favor σ phase formation, whereas Ni, Co, Al, C, and N retard its precipitation. Sigma phase precipitation in austenitic stainless steels occurs between 600 °C and 900 °C. It precipitates mainly on grain boundaries and incoherent twin boundaries, and its morphology is usually equiaxed. Sigma phase precipitation has a very slow kinetics and its formation can take hundreds and sometimes thousands of hours, see Fig. 4. There are at least three reasons for these slow kinetics: (i) C and N are insoluble in the σ phase, as a consequence, σ phase normally appears only after carbide and nitride precipitation has already taken place and the matrix is impoverished of C and N; (ii) its nucleation is difficult on account of its crystal structure being complex and very different from the parent austenite; and (iii) it is very rich in substitutional elements, thus requiring protracted diffusion times.

(b) Chi (χ) (Fe–Cr–Mo and Fe–Cr–Ti)

The χ phase is formed in the same temperature range of the σ phase (Weiss and Stickler 1972, Peckner and Bernstein 1977, Marshall 1984). Similarly to the case of M_6C carbide mentioned above, χ phase precipitation can only occur if Mo or Ti is present. Its composition is similar to that of the σ phase but, in contrast to σ phase, C can dissolve in the χ phase. Due to this property, the χ phase was classified in the past as a carbide type $M_{18}C$. The χ phase is found to form mainly on the grain boundaries, incoherent twin boundaries, coherent twin boundaries and on dislocations within the matrix. The χ phase exhibits the following orientation relationship with austenite in the case of intragranular precipitation:

$$(111)\gamma \| (110)\chi; \quad [0\ 1\ \bar{1}]\gamma \| [\bar{1}\ 1\ 0]\chi; \quad [\bar{2}11]\gamma \| [001]\chi$$

Owing to its ability to dissolve C and also to its easier nucleation, χ-phase precipitation, when it occurs, precedes σ phase formation.

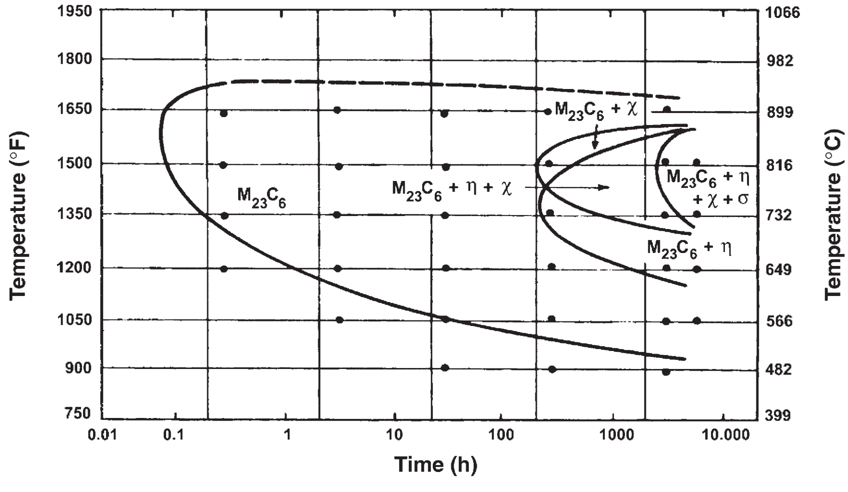


Figure 4
Time–temperature–precipitation diagram of the AISI 316 austenitic stainless steel solution annealed for 1.5 h at 1260 °C and water quenched prior to aging (after Weiss and Stickler). The stability regions of the various precipitated phases are represented by C-curves. σ = sigma phase, χ = chi phase, η = eta phase (Laves phase).

(c) Laves (Fe_2Mo, Fe_2Nb, and Fe_2Ti)

Fe_2Mo is the most common Laves phase type that precipitates in stainless steels. Nevertheless, in stabilized steels, depending on the ratio Ti:C or Nb:C, the Laves phases Fe_2Ti and Fe_2Nb can also be found (Weiss and Stickler 1972, Peckner and Bernstein 1977, Harries 1981, Marshall 1984). Its precipitation is favored by Si additions and retarded by C dissolved in the matrix. The precipitates can be found occasionally on coherent twin boundaries and grain boundaries but predominantly in the interior of the grains. The appearance of Laves phases causes some precipitation hardening and exhibits the following orientation relationship with austenite:

$$(111)\gamma \| (0001)\mathrm{Fe_2Nb}; \quad [\bar{1}10]\gamma \| [10\ \bar{1}\ 0]\mathrm{Fe_2Nb}$$

or

$$(111)\gamma \| (10\ \bar{1}3)\mathrm{Fe_2Mo}; \quad [\bar{1}2\ \bar{1}]\gamma \| [\bar{1}2\ \bar{1}0]\mathrm{Fe_2Mo}$$

(d) Other intermetallic phases

Other intermetallic phases can be observed in austenitic steels such as γ', G, and R phases (Harries 1981). For example, γ' ($Ni_3(Al,Ti)$), which is coherent with the parent austenite matrix, can be used for precipitation strengthening on some higher-Ni austenitic Fe–Cr–Ni alloys containing Ti and Al, such as Alloy 800. In another context, γ' (Ni_3Si) can be induced by neutron or charged particles irradiation at temperatures in the range 350–750 °C. Such an irradiation can also induce the precipitation of M_6C and G ($Ni_{16}Nb_6Si_7$ or $Ni_{16}Ti_6Si_7$) phase. The R phase can also be observed under some special circumstances and has been found as a consequence of aging certain austenitic weld metals at 500–900 °C.

2. *Precipitation from Low Alloy/Microalloyed Austenite*

The addition of small amounts ($\leqslant 0.1$ wt.%) of strong carbide/nitride formers elements such as Nb, Ti, V, and Zr, with Al also normally present, to the austenite (“microalloying”) is extremely important in commercial low alloy steels. Their main role is to precipitate from austenite-forming carbides/nitrides or complex carbonitrides. The solubility of these elements is small, so that high austenitizing temperatures in excess of 1100 °C are usually required. The lowest solubility product is that of TiN followed by NbC, TiC, and VN, with VC being by far the most soluble carbide(Rios, 1991). Precipitation processes from the austenite can be divided in three groups:

(i) Precipitation from undeformed austenite: this can happen after reheating of the steel or during cooling from a high soaking temperature. The carbides/nitrides formed in this way, with size ~20 nm, are intended to restrain grain growth by pinning the grain boundaries (Rios 1987).

(ii) Precipitation from deformed austenite: it is normally accepted that deformation significantly accelerates precipitation. In turn, this precipitation can delay the onset of austenite recrystallization that is invaluable for the controlled rolling of the austenite (Hansen *et al.* 1980).

(iii) Precipitation during austenite–ferrite transformation: if enough alloying element remains in solid solution after high-temperature processing, carbide/nitride precipitation can occur during the austenite–ferrite transformation. These precipitates form at the austenite–ferrite transformation front resulting in precipitates arranged in parallel planes, which is often called interphase precipitation (Honeycombe 1976). Although this sort of morphology is common in microalloyed steels, particularly those containing V, it is not restricted to them. Interphase precipitation has also been found in a variety of systems including those in which the precipitates were not carbides or nitrides such as Fe–Cu.

For the sake of completeness, one should mention that interphase precipitation is not the only way that carbides can form as a consequence of austenite–ferrite transformation. When alloy content is low and the reaction kinetics is fast, one often observes that precipitation occurs in the ferrite behind the ferrite/austenite interface. On the other hand, for higher alloying contents and slower reaction kinetics, fiber-shaped precipitates can form (Honeycombe 1976).

In practice, more than one microalloying element is often added, each for a different purpose. Hence, the low solubility of TiN makes it a good choice to restrain grain growth during reheating. Nb(C,N) is only sparingly soluble below 1000 °C, and it is a good choice if one wishes either to restrain grain growth below this temperature or more importantly to precipitate in deformed austenite. VC, on the other hand, is much more soluble and is added so that it can remain in solution and precipitate later, forming a fine hardening dispersion in the ferrite. The presence of more than one alloying element can result in the precipitation of complex multicomponent carbonitrides such as (Nb,V)(C,N) or (Nb,Ti)(C,N) (Zou and Kirkaldy 1991). The solubility of those complex carbonitrides can be found to a first approximation from the solubility of the individual carbides and nitrides (Rios, 1991).

Finally, Nb and Ti can be added to low alloy steels to form carbides and nitrides with a totally different purpose. In the “interstitial free” steels, Nb and Ti are added to act as “scavengers” for C and N. They form low-solubility precipitates that result in an almost interstitial free ferrite. This idea is actually very similar to the one used earlier in stainless steels in which MC formers were added with the purpose of combining with excess C and in this way trying to prevent $M_{23}C_6$ precipitation as previously noted.

3. Summary

Precipitation from high alloy austenite can result in a variety of carbides/nitrides/borides and intermetallic phases. The kind of precipitate formed is mainly dependent on the composition of the austenite, the temperature and the relative kinetics of precipitation of the diverse phases. One interesting point is that precipitation reactions, with the exception of precipitation hardening reactions, are normally undesirable and often lead to the degradation of some material property. In spite of this, such reactions are often unavoidable, as they are associated with processing or service conditions of the particular austenitic system. In low alloy austenite, the situation is comparatively more simple, as the most common phases to form are carbides/nitrides. Nevertheless, the engineering importance of these microalloyed systems is such that detailed knowledge is required.

Bibliography

Davis J R (ed.) 1994 *ASM Speciality Handbook®: Stainless Steels*. ASM International, Ohio

Hansen S S, Vander Sande J B, Cohen M 1980 Niobium carbonitride precipitation and austenite recrystallization in hot-rolled microalloyed steels. *Metall. Trans. A*. **11A**, 387–402

Harries D R 1981 Physical metallurgy of iron-chromium-nickel austenitic steels. *International Conference on Mechanical Behaviour and Nuclear Applications of Stainless Steel at Elevated Temperatures*. Varese, Italy, 20–22th May, Metals Society, London

Heino S, Knutson-Wedel E M, Karlsson B, Karlsson 1999 Precipitation behavior in heat affected zone of welded superaustenitic stainless steels. *Mater. Sci. Technol. (London)* **15**, 101–8

Honeycombe R W K 1976 Transformation from austenite in alloy steels. *Metall. Trans. A*. **7A**, 915–36

Jargelius-Pettersson R F A 1996 Precipitation trends in highly alloyed austenitic stainless steels. *Zeitschrift für Metallkunde* **89**, 177–83

Lee T H, Kim S J, Jung Y C 2000 Crystallographic details of precipitates in a Fe-22Cr-21Ni-6Mo(N) superaustenitic stainless steels aged at 900 °C. *Metall. Mater. Trans. A*. **31A**, 1713–23

Machado I F, Padilha A F 1996 Precipitation behavior of 25% Cr–5.5% Ni austenitic stainless steel containing 0.87% nitrogen. *Steel Res*. **67**, 285–90

Marshall P 1984 *Austenitic Stainless Steels: Microstructure and Mechanical Properties*. Elsevier Applied Science Publishers, London and New York

Padilha A F, Rios P R 2002 Decomposition of austenite in austenitic stainless steels. *ISIJ Int*. **42**, 325–37

Padilha A F, Schanz G 1980 Precipitation of a boride phase in 15% Cr-15% Ni-Mo-Ti-B austenitic stainless steel (Din 1.4970). *J. Nucl. Mater*. **95**, 229–38

Padilha A F, Plaut R L, Rios P R 2003 Annealing of cold-worked austenitic stainless steels. *ISIJ Int*. **43**, 135–43

Peckner D, Bernstein I M 1977 *Handbook of Stainless Steels*. McGraw-Hill Book Company, New York

Powell G l F, Bee J V 1996 Secondary carbide precipitation in a 18 wt.%Cr-1wt.% Mo white iron. *J. Mater. Sci*. **31**, 707–11

Rios P R 1987 A Theory for grain boundary pinning by particles, Overview no. 62. *Acta Metall*. **35**, 2805–14

Rios P R 1991 Method for the determination of mole fraction and composition of a multicomponent fcc carbonitride. *Mater. Sci. Eng. A*. **142**, 87–94

Sourmail T 2001 Precipitation in creep resistant austenitic stainless steels. *Mater. Sci. Technol*. **17**, 1–14

Tabrett C P, Sare I R, Ghomashchi 1996 Microstruture-property relationships in high chromium white iron alloys. *Int. Mater. Rev*. **41**, 59–82

Villars P, Prince A, Okamoto H 1995 *Handbook of Ternary Alloy Phase Diagrams*. ASM International, Ohio

Weiss B, Stickler R 1972 Phase instabilities during high temperature exposure of 316 austenitic stainless steel. *Metall. Trans*. **3**, 851–66

Yang K H, Choo W K 1994 The variants of the orientation relationship between austenite and cementite in Fe-30wt.%Mn-1.0wt.%C alloy. *Acta Metall. Mater*. **42**, 263–9

Zou H, Kirkaldy J S 1991 Carbonitride precipitate growth in titanium/niobium microallyed steels. *Metall. Trans. A*. **22A**, 1511–24

P. R. Rios and A. F. Padilha
Universidade Federal Fluminense,
Rio de Janeiro, Brazil

This Page Intentionally Left Blank

S

Sheet Steel: Low Carbon

Low-carbon steel sheet is available for use in both the hot-rolled condition and, after subsequent cold-rolling, annealing, and possibly coating. Modern processing provides a high degree of control of both chemistry and processing to enable a wide range of grades with different combinations of strength and formability to be produced. Close control of gauge, width, shape, and flatness is also achieved. It is the ability of low-carbon steel, however, to be formed economically and satisfactorily into a wide range of complicated shapes without splitting, necking, or wrinkling that is ensuring its continued use as a major engineering material. Other advantages include low cost, high elastic modulus, and good energy absorbing characteristics.

Sheet steels are supplied in two main categories: low-strength or mild steels for which the main requirement is a satisfactory minimum level of formability; and high-strength steels with, as expected, a minimum level of strength. Formability itself has two important aspects—drawability and stretchability. Drawability is usually characterized by the mean Langford value (often known as the *r*-value) and depends mainly on developing a suitable orientation texture containing a high proportion of (111)-oriented grains and a low proportion of (100)-oriented grains. Stretchability, usually characterized by the elongation value in a tensile test or the work hardening coefficient *n*, is mainly influenced by the strength of the steel and the strengthening mechanisms used to develop the strength. Some of the relationships between strength and elongation in a tensile test for different types of steel are illustrated in Fig. 1 while equivalent relationships between strength and drawability (mean *r*-value), are illustrated in Fig. 2. These figures show that there is a general reduction in formability with increasing strength and that steels strengthened by different mechanisms are able to develop various ranges of strength to provide different combinations of strength and formability.

1. General Processing Considerations

Almost all sheet steel is now made using a continuous casting process to produce slabs 200–250 mm thick. After cooling these slabs are usually reheated to temperatures up to 1250 °C for hot rolling. The first stage of hot rolling, called roughing, reduces the thickness usually into the range 30–45 mm and the second stage, called finishing, reduces the thickness to the final hot-rolled gauge required, often in the range 1–2 mm up to 5–12 mm. The steel leaves the last stand of the hot mill at a finishing temperature which is determined by metallurgical requirements. Water cooling is then used along a runout table to achieve the required temperature for coiling. Over the last few years thin-slab casting processes have been introduced in a few steel plants which enable intermediate gauges, usually obtained by roughing, to be cast directly.

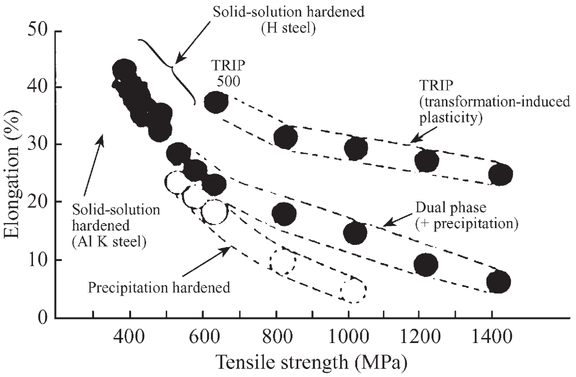


Figure 1
Strength–ductility balance of high-strength steel (after Shimada *et al.* 1996).

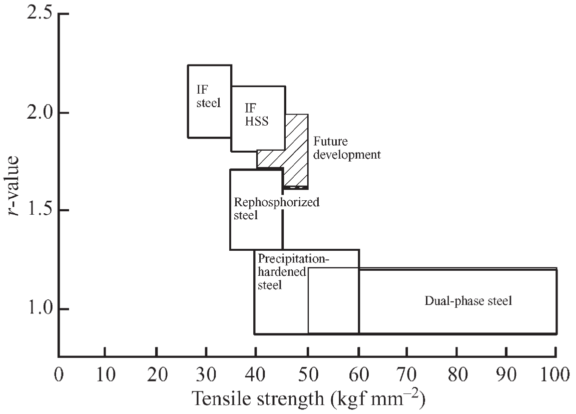


Figure 2
Relationship between mean *r*-value and tensile strength (after Takechi 1990).

Cold rolling is used for gauges usually in the range 0.4–3 mm using cold reductions of 50–80%. During cold rolling the steel becomes hard and loses its ductility; therefore an annealing process is used to restore ductility and to develop almost the final properties. In the batch process several tight coils are stacked on a furnace base and heated and cooled under a protective atmosphere, the complete process usually taking several days. In the continuous process each coil is uncoiled and then recoiled after passing through a furnace, the complete process taking several minutes. After annealing, cold-rolled steels may be given a light cold reduction, called temper rolling, primarily to remove the yield point but also to improve flatness and develop the required surface roughness and texture. Cold-rolled and annealed steel would normally have a better formability, gauge control, flatness, and surface than a hot-rolled steel but the former would have a higher cost due to the additional processing.

2. *Mild Steel*

2.1 Hot-rolled Mild Steel

The most formable standard grade of hot-rolled mild steel often contains 0.03–0.06% carbon, 0.003–0.006% nitrogen, 0.2–0.3% manganese, and 0.03–0.06% aluminum and is processed using a moderately high coiling temperature (~650 °C). An austenite with a relatively coarse grain size is produced at the slab reheat stage, which, during hot rolling, recrystallizes between each deformation pass to give a smaller grain size. A moderately fine grain size is finally produced after transformation to ferrite depending on the time to the onset of water cooling, the cooling rate (which depends on thickness), and the coiling temperature. The orientation texture is usually fairly random with the result that mean r-values are low and close to unity. The drawability is, therefore, moderate. Lower formability and slightly higher-strength grades are produced by relaxing the upper carbon and manganese limits, whereas a superior formability grade may be produced by combining the interstitial nitrogen with a small addition of titanium or boron.

2.2 Cold-rolled Mild Steel

There are two main types of cold-rolled mild steel. The first is the more traditional type described as aluminum-killed (AK) steel, in which the carbon is relatively free in the structure and may form iron carbides. The second, called interstitial free (IF) steel, contains a strong carbide-forming element such as titanium and/or niobium to combine with some or all of the carbon and nitrogen. IF steels of today usually contain ultralow levels of carbon and nitrogen (both <0.003%), whereas the AK mild steels have carbon contents above 0.015% and often higher, depending on the properties required and the annealing method to be used. IF steels often have mean r-values above 2.0 and are used for the most difficult forming applications. The highest r-values of AK steels are usually a little below 2.0 but grades with mean r-values as low as 1.0 may be obtained depending mainly on the composition and annealing method used. These grades are clearly used for applications requiring modest formability.

(a) Interstitial free mild steels

The IF steels produced in the early 1970s generally contained about 0.01% carbon and needed more than a stoichiometric ratio of titanium to carbon, nitrogen, and sulfur to give the high mean r-values. The use of the ultralow carbon contents indicated above has enabled reasonably high r-values to be obtained with the addition of less than a stoichiometric amount of titanium, but higher r-values can still be obtained using more than a stoichiometric addition. When niobium is used, separately or in combination with titanium, there is evidence that optimum r-values and values of elongation are obtained with niobium: carbon atomic ratios of less than unity.

The use of low slab reheat temperatures and high coiling temperatures with IF steels leads to superior forming properties, but a finer hot-band grain size also leads to higher r-values. Such a grain size may be obtained by cooling rapidly as soon as possible after the steel emerges from the last stand of the hot mill, commonly referred to as "early cooling."

All IF steels develop the highest r-values when high cold reductions, such as 80%, are used. Similar compositions and prior processing may be used for both batch or continuous annealing and similar properties are obtained, but most IF steels are continuously annealed. In general, the annealing requirements of IF steels are not critical and this makes them particularly suitable for processing on hot dip coating lines if a very formable product is required. Properties improve, however, with increasing annealing temperature and temperatures up to 850 °C are often used. IF steels do not normally exhibit a yield point. They require, therefore, only the minimum of temper rolling to improve shape and impart the required surface roughness. Many IF steels are also completely non-aging unless specific measures are taken to develop bake hardening (see below).

(b) Aluminum-killed mild steel

Aluminum-killed mild steels may also be processed by batch or continuous annealing. The metallurgical requirements are, however, completely different for the two processes if optimum formability is to be obtained. With the older batch annealing method, the most important requirement is that aluminum and nitrogen must be held in solution in the hot-rolled coil. Aluminum nitride is then able to cluster or precipitate during slow heating to influence the recrystallization process to give high r-values. Retaining aluminum nitride in solution is achieved by the use of a high slab reheat temperature close to 1250 °C and a low coiling temperature, usually below 600 °C.

Optimum properties are usually obtained with 0.003–0.006% nitrogen and 0.025–0.04% soluble aluminum. Carbon contents are usually close to 0.03–0.04%, though higher carbon contents may be used for lower formability grades. With all of these carbon contents, the carbon that is taken into solution at the annealing temperature is able to reprecipitate almost completely during the long slow cool to room temperature. The steel then becomes completely nonaging. Cold reductions close to 70% are often used since higher and lower cold reductions tend to give lower *r*-values. The final temper rolling must be sufficient to remove the yield point. About 0.8% reduction in thickness is often used.

There are two important requirements for achieving optimum properties in AK steel by continuous annealing. First, the nitrogen must be combined as aluminum nitride in the hot-rolled coil to give low interstitial nitrogen contents and second, the carbides must be coarse. The coarse carbides dissolve slowly during the rapid heating. Recrystallization is then able to take place in a substantially interstitial-free matrix (before the carbides have dissolved) to give an adequately coarse grain size and a beneficial texture, to give relatively high *r*-values. The presence of aluminum nitride and a low interstitial nitrogen content in the hot band is promoted by the use of a low slab reheat temperature (down to 1100 °C and below) and sometimes a high aluminum content (up to 0.09%). Both of these factors minimize the dissolution during slab reheating of the aluminum nitride formed previously. The most important requirements are, however, that the total nitrogen content is low ($<0.003\%$) and that the coiling temperature is high, well above 700 °C. This enables aluminum nitride to form as the steel cools below the coiling temperature. The coarse carbides are also obtained by the use of the high coiling temperature.

The properties of continuously annealed AK steels benefit from high cold reductions and high annealing temperatures like IF steels, but unlike IF steels the final properties depend critically on the cooling part of the annealing cycle. This must be designed to reprecipitate the carbon that goes into solution at the annealing temperature as completely as possible. To achieve this, the steel is first cooled slowly at about $10\,^\circ\mathrm{C\,s^{-1}}$ to approximately 675 °C to maximize the carbon in solution. This is followed by rapid cooling, ideally at about $100\,^\circ\mathrm{C\,s^{-1}}$, to a temperature usually below 450 °C. This minimizes the amount of carbon reprecipitated during cooling to maximize the driving force for the nucleation of fine closely-spaced carbides within the grains. These are then able to grow by short-range diffusion. If cooling to about 400 °C or above is used the steel is allowed to cool slowly to about 350 °C over 2–3 minutes to allow the carbon to precipitate. If temperatures close to 250 °C or below are used, the steel is reheated to about 350 °C or above and then allowed to cool slowly over approximately two minutes before rapid cooling to room temperature. Cooling to 250 °C instead of 400 °C leads to the nucleation of more closely spaced carbides and the use of reheating enables them to grow rapidly by taking advantage of a higher diffusion rate. The reprecipitation of carbon then becomes more complete and the steel is less likely to be subject to strain aging. Carbon contents close to 0.01% lead to poor reprecipitation of carbon even with a potentially good cooling sequence. To give the best possible properties, carbon contents tend, therefore, to be in the range 0.015 to 0.025%. The use of relatively low manganese contents such as 0.15% can lead to the formation of relatively more fine manganese sulfide precipitates which may act as easy sites for the nucleation of new carbide precipitates during cooling. A low manganese content is therefore beneficial.

The use of a high coiling temperature can lead to operational problems. An addition of boron may be used to form boron nitride, thereby providing an alternative method of reducing the hot-band interstitial nitrogen content without the need for a high coiling temperature. The benefit to be obtained from coarse carbides is, however, lost with the result that this type of steel develops low *r*-values. Therefore it may only be used for applications requiring relatively low formability. All continuously annealed AK steels exhibit a yield point as in the annealed condition. They all, therefore, require about 1% temper rolling.

3. High-strength Steels

3.1 Solid Solution-strengthened Steels

Each type of cold-rolled and annealed mild steel may be given a modest increase in strength, to give yield stresses usually in the range 220–300 MPa, by the addition of certain elements that remain in solution in the final product. The element most commonly used is phosphorus because it gives a worthwhile increase in strength for small additions but the strength increase may be supplemented by additions of manganese or silicon. Boron may also be used in titanium IF steels. The phosphorus addition is usually restricted to below 0.1% to avoid problems with welding and other potential problems. The main advantage of these steels is that they retain *r*-values almost as high as those of the mild steels from which they are derived There is, however, the usual progressive drop in formability with increasing strength even though this tends to be less than for other strengthening mechanisms. There is some evidence for a niobium IF steel that the decrease in *r*-value for a given increase in strength is less for phosphorus than for manganese or silicon, whereas for a titanium IF steel the decrease in elongation is greater for phosphorus than for the other two elements.

The base compositions of all these steels are usually similar to those of the mild steels from which they are derived and the general processing requirements are also the same. Thus, for example, a continuously annealed solid solution-strengthened AK steel must be hot-rolled to precipitate aluminum nitride in the hot band if high *r*-values are to be obtained. Similarly, a batch-annealed AK steel must be hot-rolled to retain nitrogen in solution.

Additions of manganese or silicon may be used to give small solid solution-strengthening effects in hot-rolled steels but these additions are more usually used in hot-rolled steels to influence transformation characteristics.

3.2 Bake-hardening Steels

A bake-hardening steel is one that increases its yield stress during a paint stoving operation after forming by means of a strain-aging process. In the as-delivered condition, therefore, the steel has the formability of a relatively low-strength steel but the performance in service of a steel with a higher yield stress. The bake-hardening tendency is usually measured by straining a steel by 2% and then baking at 170 °C for 20 minutes. The difference between the flow stress after 2% strain and the lower yield stress after baking is taken as a measure of the bake hardening. The minimum levels of bake hardening that are usually considered to provide a useful benefit are 30–40 MPa and maximum values usually used are 50–60 MPa. Higher values than these would lead to excessive strain aging at room temperature with a consequent loss in formability and the formation of stretcher strain markings on forming, unless just-in-time delivery is organized before forming.

Any steel that exhibits strain-aging will automatically exhibit bake hardening but the commonly used bake-hardening steels are based on solid solution-strengthened steels. Their chemistry or processing is modified, if necessary, to leave sufficient carbon in solution after annealing to give the bake hardening. It is usually considered that 5–20 ppm carbon in solution is adequate to give sufficient bake hardening but different values in this range are preferred by different suppliers depending on other details of the steel. Obtaining a suitable solute carbon content is a natural consequence of the continuous annealing of an AK solid solution-strengthened steel, since even an optimized cooling path is not able to reprecipitate all the carbon completely. A batch-annealed AK bake-hardening steel may be obtained by reducing the carbon content of the steel to about 0.01% or below. The complete reprecipitation of carbon during slow cooling then becomes difficult and sufficient carbon remains in solution to give the bake hardening even after the long slow cool. Batch annealing using hydrogen as the protective atmosphere gives more bake hardening than the older HNX gas annealing due to the faster cooling rate with hydrogen.

Continuously annealed bake-hardening IF steels, containing ultralow carbon contents, may be produced in several ways depending on the *r*-values required but most of these steels are alloyed with niobium and/or titanium. With the lower alloy additions, some of the carbon remains uncombined and this carbon is available to provide the bake hardening. With higher additions, most of the carbon is combined with titanium or niobium prior to annealing. The free carbon needed to provide the bake hardening is then obtained during annealing by the use of a high annealing temperature such as 850 °C or more, to take carbon back into solution. The steel is then cooled rapidly to minimize the reformation of precipitates on cooling. As an example, a minimum cooling rate of $30\,^{\circ}\mathrm{C\,s^{-1}}$ from 850 °C is necessary to give adequate bake hardening for a steel with a niobium:carbon atomic ratio of unity. With a lower niobium to carbon atomic ratio, such as 0.3, the very high annealing temperature and the rapid cooling would not be necessary but the resulting *r*-values would be lower.

The production of bake-hardening IF steels by batch annealing is only possible using relatively low alloy additions, since annealing temperatures approaching 850 °C are not possible with this annealing method. Clearly obtaining a consistent level of bake hardening becomes difficult when it depends on a very precise control of steel chemistry.

3.3 Microalloyed or High-strength Low-alloy (HSLA) Steels

A useful method of increasing the strength of a hot-rolled steel is to increase the carbon and manganese contents. The potential strength increase is, however, limited if the content of these elements is restricted to retain weldability, unless hard transformation products are obtained. A larger increase in strength may be obtained by making small additions of niobium, titanium, or vanadium combined with a suitable choice of carbon, manganese, and nitrogen content. These steels, called microalloyed or high-strength low-alloy steels (HSLA steels), develop their strength mainly by grain refinement, different amounts of precipitation-strengthening, and a little solid solution-strengthening. They usually have yield stresses in the range of 300–500 MPa or more, but the most commonly used grades have yield stresses a little below the middle of this range.

The grain refinement in these microalloyed steels arises from several mechanisms. At the slab reheat stage, residual precipitates restrict grain growth. At lower temperatures the recrystallization of austenite between each rolling pass is restricted, mainly by strain-induced precipitation but also by solute drag effects. Greater deformation is then accumulated in the austenite before recrystallization can take place and this leads automatically to a finer austenite grain size. Finally, at sufficiently low temperatures, recrystallization is completely inhibited. The final transformation to ferrite then occurs from an unrecrystallized austenite, which again leads to a finer ferrite grain size. It is generally considered that of the three elements mentioned the greatest retardation of recrystallization for a given addition occurs with niobium and the least with vanadium. Niobium tends, therefore, to be the most commonly used element for medium-to-large increases in strength but the effect of niobium may be supplemented by vanadium to provide additional precipitation-strengthening from vanadium nitride if a very high yield stress (~ 500 MPa) is required.

Cold-rolled versions of microalloyed steels may be obtained using either batch or continuous annealing but for a given composition higher strengths are usually obtained by continuous annealing. A continuously annealed product also has the potential for giving more consistent properties. In general the strength range possible with a cold-rolled and annealed steel is lower than that for a hot-rolled steel and it is useful to note that all HSLA steels have low *r*-values. They cannot, therefore, be used for applications requiring high drawability.

3.4 *Transformation-strengthened Steels*

A further method of producing higher-strength steels is to develop a significant proportion (e.g., up to 20%) of hard transformation products in a matrix of fine ferrite grains. The earliest of these steels to be developed were called dual-phase steels in which the second phase was martensite. These steels may in principle be obtained directly from the hot mill in two ways and by continuous annealing. In the first direct method, a carbon–manganese steel is used containing sufficient manganese and other additions (such as chromium) to suppress the formation of bainite during runout table cooling. The steel develops a suitable structure of ferrite and high-carbon austenite during the early stages of this cooling and the austenite finally transforms to martensite after coiling provided that the coiling temperature is below the M_S (martensite start) temperature ($\sim 250\,°\mathrm{C}$). A difficulty with this method is that not many hot mills are able easily to achieve the required very low coiling temperature. In the second direct method, the steel is more highly alloyed with manganese, silicon, and molybdenum to suppress the formation of bainite even when a coiling temperature well above the M_S temperature is used. A difficulty with this method, however, is that the relatively high silicon level often leads to a poor surface in any hot-rolled product. Neither method is, therefore, in common use.

Dual-phase ferrite–martensite steels are more easily made after cold rolling by continuous annealing. A suitable ferrite–high-carbon austenite microstructure is developed by intercritical annealing and initial slow cooling. The usual rapid cooling available on a continuous annealing line enables the steel to be cooled to below the M_S temperature without the formation of bainite and the austenite then transforms to martensite.

More recently, a hot-rolled product has been developed involving a final structure of ferrite and up to about 20% of bainite. A carbon–manganese steel is hot rolled using a low finishing temperature but still in the austenite region, and a suitable ferrite–high-carbon austenite is developed on the runout table either by continuous cooling or by interrupted cooling part way along the runout table. The use of a coiling temperature usually in the range 400–450 °C enables the austenite that is retained to transform to bainite. This type of product is finding widespread application in structural parts. An equivalent cold-rolled product could be made by continuous annealing.

A further type of steel has been developed in which the hard phase may be a mixture of retained austenite and bainite. These steels may easily have a tensile strength above 600–800 MPa, depending on the precise composition used, and are called multiphase or TRIP steels (transformation-induced plasticity), because the retained austenite transforms to martensite under the action of the forming strains required to make a component. This leads to the good formability already noted from Fig. 1. The steels needed are based on a carbon–manganese composition but with a high level of silicon or possibly aluminum to suppress carbide precipitation. Processing on the hot mill is similar to that for the ferrite–bainite steels mentioned above. In this case, however, the gradual formation of bainite after coiling leads to a further buildup of carbon in the austenite yet to transform. The result is that the M_S temperature of this austenite reduces to below room temperature. The austenite is therefore retained in the structure after final cooling. A difficulty with this type of hot-rolled steel, however, is that it requires extremely close control of hot-mill processing if acceptably consistent properties are to be obtained, and so it is not yet in commercial production. These types of steel may also in principle be processed by continuous annealing for which both the temperature and time for the partial transformation of austenite to bainite after rapid cooling would be under closer control. In this case certain research has shown that a high holding time is beneficial because the final austenite retained at room temperature transforms gradually under the action of forming strains, thereby imparting the highest elongation. This is because austenite regions with different carbon contents have different stabilities in the presence of deformation.

Ultrahigh-strength steels are steels with a tensile strength above about 1000 MPa, though strengths above 800–900 MPa could also be regarded as ultrahigh strength. Two types are possible, depending on whether the structure consists primarily of martensite or of bainite. However, each may contain a proportion of the other phase together with ferrite or retained austenite and may include precipitation effects. These steels may be processed by continuous annealing involving rapid cooling, using compositions usually based on carbon, manganese, and silicon.

See also: Austenite Decomposition: Overall Kinetics during Isothermal, and Continuous Cooling Transformation

Bibliography

Asfahani R, Tither G (eds.) 1993 *Proc. Int. Symp. Low Carbon Steel for the 90s*. TMS, Pittsburgh, PA

Bleck W (ed.) 1998 *Proc. Int. Symp. Modern LC and ULC Sheet Steels for Cold Forming: Processing and Properties*. Verlag Mainz, Aachen, Germany

Bramfitt B L, Mangonon P L (eds.) 1982 *Proc. Symp. on Metallurgy of Continuous Annealed Sheet Steel, Dallas*. AIME, Warrendale, PA

Collins L E, Baragor D L (eds.) 1991 *Proc. Conf. on Interstitial Free Sheet: Processing, Fabrication and Properties*. CIM/ICM, Ontario

Cuddy L J 1975 *Plastic Deformation of Metals*. Academic Press, New York

Iron and Steel Institute of Japan (ed.) 1994 *Int. Forum Physical Metallurgy of IF Steel* 1994. ISIJ, Tokyo

Korchynsky M (ed.) 1977 *Microalloying '75*. Union Carbide Corporation, New York

Kot R A, Bramfitt B L (eds.) 1981 *Proc. Conf. Fundamentals of Dual Phase Steel*. AIME, Warrendale, PA

Iron and Steel Society 1995 *Microalloying '95, Conf. Proc.* The Iron and Steel Society, Pittsburgh, PA

Pickering F B (ed.) 1992 Materials Science and Technology. 1992 VCH, Vol. 7, Chaps. 6 and 7

Pradhan R (ed.) 1984 *Proc. Symp. on Metallurgy of Continuously Annealed Cold Rolled Sheet Steel, Detroit*. AIME, Warrendale, PA

Pradhan R (ed.) 1990 *Proc. Conf. Metallurgy of Vacuum Degassed Steel Products*. The Minerals, Metals and Materials Society, Warrendale, PA

Pradhan R (ed.) 1992 *Proc. Conf. on Developments in the Annealing of Sheet Steels*. The Minerals, Metals and Materials Society, Warrendale, PA

Pradhan R (ed.) 1994 *Proc. Symp. on High-Strength Sheet Steels for the Motor Industry*. The Iron and Steel Society, Baltimore, MD

Pradhan R, Ludkovsky G (eds.) 1988 *Proc. Symp. on Hot and Cold Rolled Sheet Steels*. The Metallurgical Society, Cincinnati, OH

Shimada M, Nashiwa M, Osaki T, Furukawa Y, Shimatani Y, Fujita Y, Aihara H 1996 *Review of Annealing Technology*. Technical Exchange Session, IISI p. 17

Takechi H 1990 In: Dorgham M A (ed.) *Vehicle Design and Components—Materials Innovation and Automotive Technology, 5th IAVD Conf. Geneva*. Interscience, New York

R. C. Hudd
Porthcawl, UK

Shell: Properties

Mollusk shells are exquisitely designed biocomposites (*biological* ceramics) consisting of calcitic or aragonitic $CaCO_3$ and a small quantity of organic components (the nonmineral portion of mineralized tissue is referred to as the *matrix* in biological jargon). They are very heavily mineralized hard tissue, the organic components comprising as little as 1% of the shell volume. According to the morphology and organization of the biominerals, mollusk shells can be divided into five common shell microarchitectures: crossed-lamellar, nacreous, prismatic, foliated, and homogeneous (Bøggild 1930). The mechanical properties of shells with different structures vary considerably (Jackson *et al.* 1988; Kuhn-Spearing *et al.* 1996). Compared with engineered ceramics or nonbiogenic (mineral) calcium carbonate, mollusk shells exhibit an extraordinary work of fracture, several orders of magnitude higher than ceramists have been able to achieve in ceramic matrix composites. Such impressive properties originate from their exquisite microstructural design, essentially the organization of the biominerals and the protein matrix.

1. Microstructural Design of Mollusk Shells

Mollusk shells invariably consist of a very thin uncalcified layer, called the *periostracum*, which covers the outer surface of the shell and an inner, much thicker calcified layer. In the calcified inner layer, the mineral phase constitutes well over 95% of the shell mass, the remainder being the organic matrix. The morphology and organization of the basic building blocks for each shell structure are different. The common characteristics of shell microstructures are that the building blocks are very small, having characteristic lengths of tens of nanometers to microns, and are organized in an ordered arrangement in at least one dimension (Gregoire 1972). The organic matrix is present as a thin envelope or sheet surrounding each mineral unit (Gregoire 1972). Here, we describe three common structures: crossed-lamellar, nacreous, and prismatic.

1.1 Crossed-lamellar Microstructure

The crossed-lamellar microstructure is the most common structure in mollusk shells, being represented in ~90% of gastropods and ~60% of bivalves and is considered to be at the pinnacle of mollusk evolution. The mineral phase is generally aragonite (although there are a few examples of calcitic crossed-lamellar structures). The morphology of the basic mineral unit is approximately rectangular or lathe-like. Structure is present in the lamellar microarchitecture at five different length scales, as will now be described.

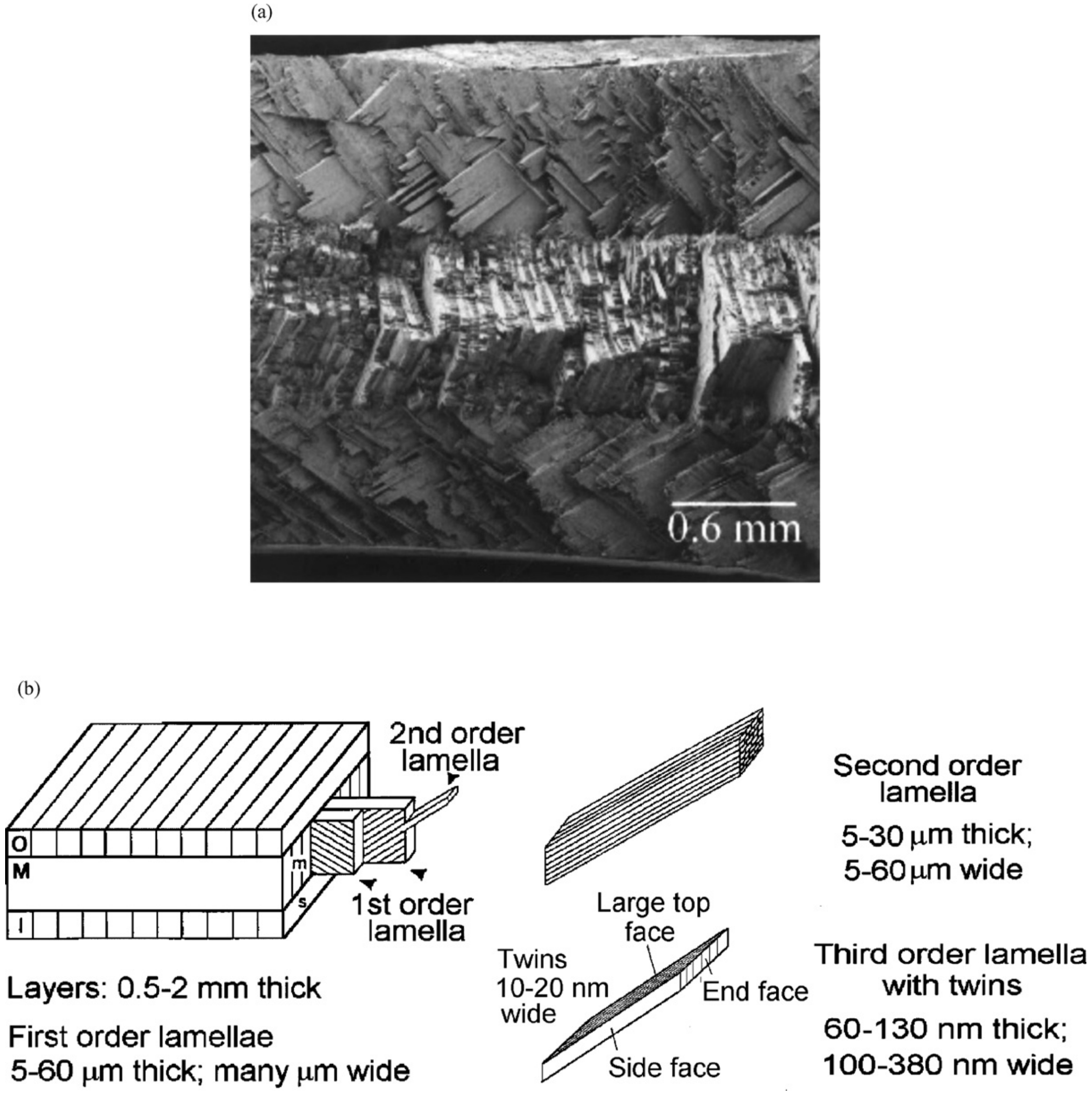

(a)
0.6 mm
(b)
2nd order lamella
O
M
I
m
s
1st order lamella
Layers: 0.5-2 mm thick
First order lamellae
5-60 μm thick; many μm wide
Large top face
Twins
10-20 nm wide
End face
Side face
Second order lamella
5-30 μm thick;
5-60 μm wide
Third order lamella with twins
60-130 nm thick;
100-380 nm wide

Figure 1
(a) SEM image of the fracture surface of a bend specimen of the shell of *Strombus gigas*. (b) Schematic drawing of the crossed-lamellar microarchitecture, including characteristic dimensions of the three lamellar orders. (O, M, and I refer to outer, middle, and inner (adjacent to the animal) layers, respectively.)

Figure 1(a) is a SEM image of a fracture surface showing a cross-section of the crossed-lamellar structure taken from a specimen of *Strombus gigas*, the pink queen conch native to Caribbean habitats; its microstructure is depicted schematically in Fig. 1(b). The five length scales present are the macroscopic layers, the first-, second-, and third-order lamellae, and twins within each third-order lamella. The basic building blocks are the third-order lamellae. Many parallel third-order lamellae form one second-order lamella, and many parallel second-order lamellae form one first-order lamella. Second-order lamellae are rotated by ~90° from neighboring first-order lamellae; the orientations of first-order lamellae also change by ~90° from one macroscopic layer to its neighboring layer.

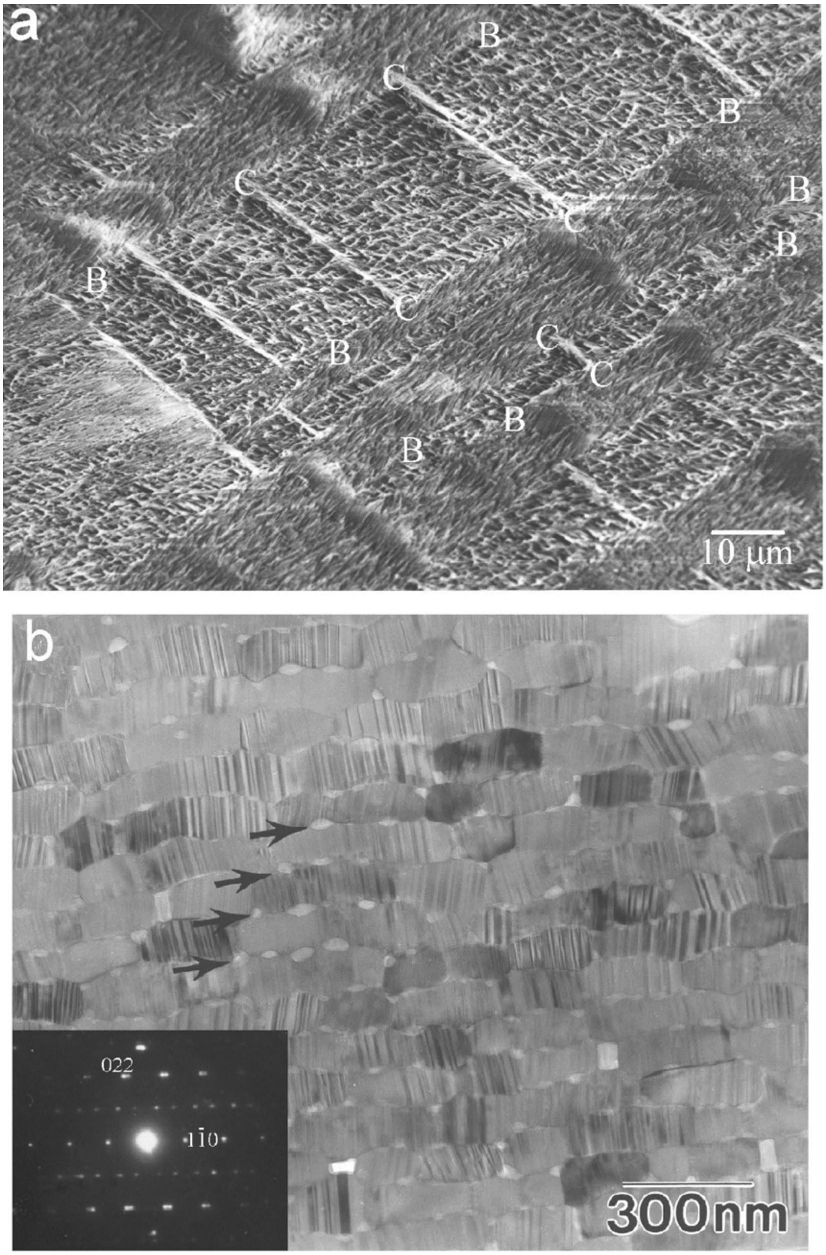


Figure 2
(a) SEM image of a polished, partially demineralized shell section from *Strombus gigas* showing interfaces separating second-order lamellae (CC) and interfaces separating first-order lamellae (BCCB and BCB). The long axes of the third-order lamellae are at an angle of about 45° to this polished section. These images confirm previous claims that the organic matrix surrounds every crystallite and is present at every interface. However, the boundaries of the second-order lamellae are thicker than the boundaries separating first-order lamellae. Within any second-order lamella, protein matrix associated with third-order lamellae can also be seen. (b) TEM ‘‘end-face’’ image of an ion-thinned specimen from the shell of *Strombus gigas*. The protein sheaths present between third-order lamellae are made up of thin membrane-like organic sheets connected to thicker globular proteins. The globular features are arranged along the top and bottom of each third-order lamella but not along the sides of these lamellae. The inset diffraction pattern reveals that the vertical features present in every third-order lamella are ‘‘polysynthetic’’ growth twins.

Shells with the crossed-lamellar structure contain only ~1 vol.% organic material, which is distributed at the interfaces between first-order lamellae, the interfaces between second-order lamellae and the interfaces between third-order lamellae. Figure 2(a) is a SEM micrograph of a polished and partially demineralized sample from *Strombus gigas*. C-C in the micrograph are the boundaries separating adjacent second-order lamellae within individual first-order lamellae, while the boundaries BCCB and BCB separate adjacent first-order lamellae; the second-order lamellae interfaces are thicker than the interfaces separating first-order lamellae. Figure 2(b) is a TEM micrograph showing the third-order lamellae "end-on." The most obvious features of the third-order interfaces are the globular matrix "particles" distributed along the largest face of the third-order lamellae. In fact, membrane-like matrix is continuous around each third-order lamella.

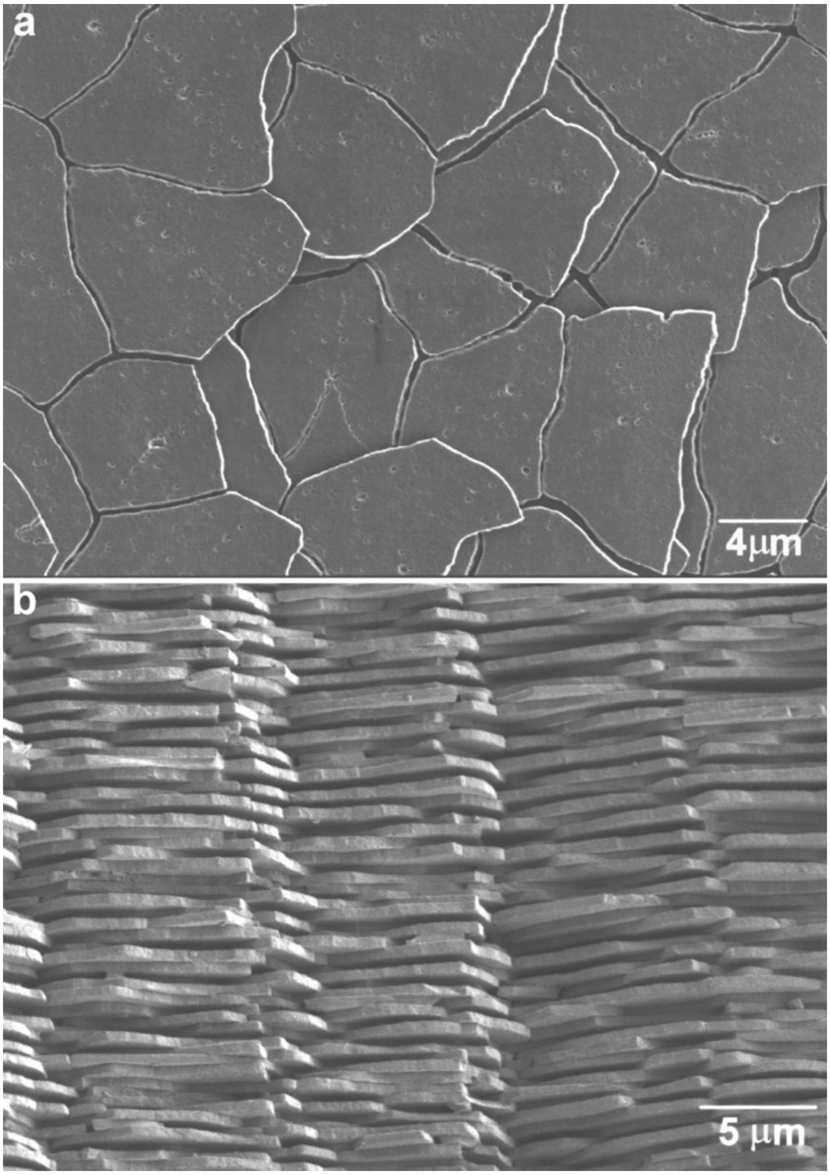


Figure 3
SEM images of the shell of red abalone, *Haliotis rufescens*. (a) From a polished plan view section, and (b) from a fracture surface.

1.2 Nacreous Microstructure

The nacreous microstructure consists of aragonite tablets, ~0.4 μm thick and 5–10 μm wide (Gregoire 1972, Laraia and Heuer 1990, Sarikaya and Aksay 1992). The tablets have an irregular form, as shown in Fig. 3(a). This example is taken from the red abalone, *Haliotis rufescens*; the shell of this gastropod has an inner aragonitic nacreous sublayer abutting onto an outer, calcitic prismatic sublayer (see Sect. 1.3). The *c* axes of the aragonite tablets are aligned and perpendicular to the inner surface of the shell. In gastropods and cephalopods, the tablets are arranged in stacks, as shown in Fig. 3(b); within each stack, all the tablets have the same crystallographic orientation. In gastropods such as the red abalone, evidence now exists (Schäffer *et al.* 1997) that the crystallographic continuity with the stacks is achieved by tablets growing through pores within the organic matrix. In bivalves, the tablets are arranged in a "brick-wall" type of configuration. The crystals in each layer develop at offset positions from those of the layer immediately below, and there is no crystallographic continuity between any two neighboring tablets.

The nacreous microstructure generally contains ~5% organic matrix, which is mainly distributed at the interfaces between the aragonite tablets as 0.3–0.5 μm thick sheets, though a small amount of matrix may be occluded within the mineral. The configuration of the organic components in nacre consists of thin layers of β-chitin sandwiched between two layers of silk fibroin-like proteins, onto which acidic macromolecules have been adsorbed (Nakahara *et al.* 1982; Nakahara 1983; Weiner *et al.* 1983).

1.3 Prismatic Microstructure

Figure 4 is a SEM micrograph of the prismatic structure at the growing edge of the shell of *Haliotis rufescens*. This structure consists of aggregates of prisms, the orientation of the prisms being oblique to the inner shell surface. Within one aggregate, all the prisms have the same orientation; their orientations vary considerably from one to another aggregate. Although, as already mentioned, the prismatic layer in *Haliotis rufescens* is calcitic, numerous aragonitic examples are also known. Individual prisms are up to 200 μm across and may be several millimeters long. Up to 5 μm thick organic matrix sheets surround each prism.

2. Mechanical Properties of Mollusk Shells

Shell specimens of the three microstructural types just described have similar properties: the elastic moduli are between 30 GPa and 60 GPa, the bend strengths are modest, 50–250 MPa, and the work of fracture is extraordinarily high, up to $\sim 13 \times 10^3$ Jm^{-2}, compared to the fracture surface energy of ~1 Jm^{-2} for

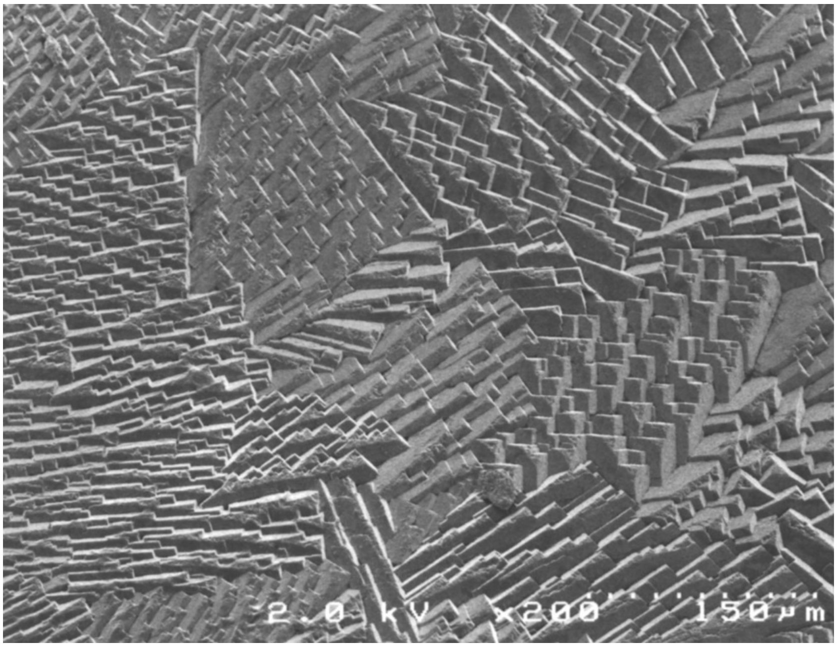


Figure 4
SEM image of the prismatic structure from the growing edge of the shell of *Haliotis rufescens*.

nonbiogenic $CaCO_3$ (Laraia and Heuer 1989, Jackson *et al.* 1990, Kuhn-Spearing *et al.* 1996, Kamat *et al.* 2000).

Young's modulus of mineral aragonite is ~100 GPa; direct measurement of the modulus of biogenic aragonite using load–deflection curves of bend specimens taken from the shell of *Strombus gigas* yielded a value of ~40 GPa. The reduction in the modulus from ~100 GPa to ~40 GPa upon addition of a small quantity of a low-modulus phase (the matrix) can be rationalized using classical laminate theory (Kamat 2000). Moreover, the crossed lamellar structure renders this shell globally isotropic.

The modest strengths noted above were calculated using classical beam theory, from the maximum load supported by rectangular specimens loaded in bending prior to catastrophic fracture. There are differences between specimens tested in the wet compared to the dry state, and the strengths that have been determined depend on how the specimens were machined out of the parent shell, due to the complexity of the microarchitecture. For these reasons, and because fracture strength is never an intrinsic property of brittle materials, it is more instructive to discuss the extraordinary work of fracture of these biological ceramics.

We begin with the crossed lamellar structure; Fig. 5(a) shows a typical load–displacement curve obtained during fracture of a "dry" specimen from the shell of *Strombus gigas* (see Kuhn-Spearing *et al.* 1996). Each serration in the load–displacement curve corresponds to a microcrack which nucleates and propagates along a first-order lamellar interface until it reaches the interface between adjacent layers, where

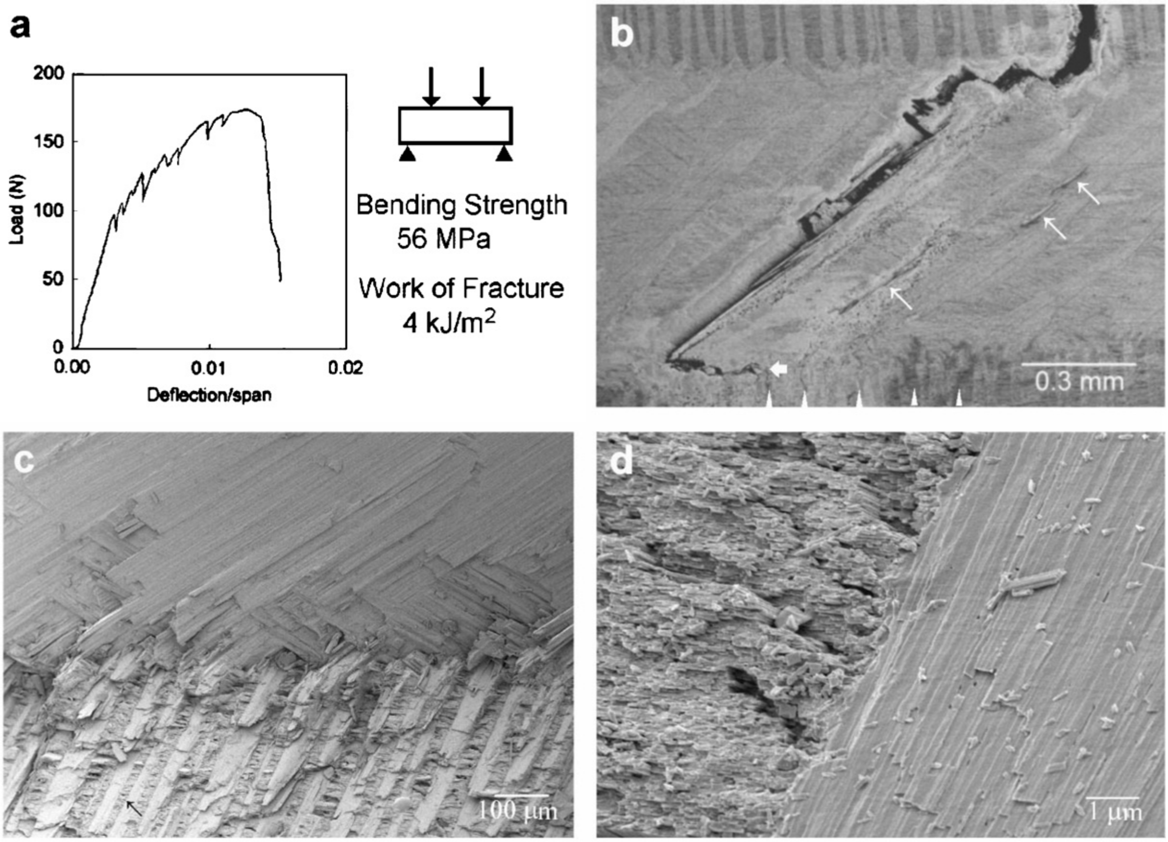


Figure 5
(a) Load–deflection curve of bend specimen of the shell of *Strombus gigas*. (b) Polished side face of a fractured bend specimen taken from the shell of *Strombus gigas*. The main strength-defining crack is apparent. The broad white arrow shows an interlayer crack; the white narrow arrows denote arrested microcracks, and the bare white arrowheads, "channel" cracks along first-order lamellar interfaces. (c) SEM image of the fracture surface of a specimen taken from the shell of *Strombus gigas*, showing the inner layer (top half) and middle layer (bottom half) of the structure. (d) Within the inner layer, the SEM image shows that alternate first-order lamellae show either rough or smooth fracture surfaces, as discussed in the text.

crack arrest occurs because of the changing orientation and increased effective toughness of the middle layer (these arrested channel cracks are indicated by the white arrowheads at the bottom of Fig. 5(b)). This "orientational" toughening arises because the second-order lamellar interfaces in alternate first-order lamellae in the middle layer are oriented such that a propagating crack cannot penetrate a weak second-order mineral/protein lamellar interface, but must penetrate and fracture the aragonite itself. Arrested microcracks also appear to be present in the middle layer, isolated from the channel cracks in the inner layer. These latter were generated during propagation of the final strength-defining macrocrack and also contribute to the total macroscopic work of fracture. Catastrophic failure only occurs when one of the channel cracks is able to penetrate the middle layer (Fig. 5(b)) and ultimately emerges from the bend specimen. However, these cracks do not propagate catastrophically through the middle layer; instead they are retarded by the bridging action of the first-order lamellae. This bridging mechanism is responsible for the largest fraction of energy dissipation during fracture of the shell. The exceptional work of fracture of mollusk shells—up to $13 \times 10^3\,\mathrm{J\,m^{-2}}$—compared with conventional brittle ceramics or mineral aragonite—of the order of $1\,\mathrm{J\,m^{-2}}$—is readily understood from load–deflection curves of the type shown in Fig. 5(a); the overall strain to failure is very large because of the ability of the shell to sustain a multiplicity of microcracks and provide propagation-resisting bridging forces on the surfaces of dominant cracks prior to catastrophic failure (Kamat *et al.* 2000).

This orientational toughening is obvious from fractographs of the middle layer of a fractured specimen. A low-magnification SEM image (Fig. 5(c)) reveals that the outer layer fracture surface is relatively smooth (the upper portion of Fig. 5(c)), due to the crack propagating along a first-order interface, whereas in the middle layer, alternate first-order lamellar fracture surfaces are either rough or smooth (the lower portion of Fig. 5(c) and at higher magnification in Fig. 5(d)). In fact, the SEM images suggest that the conch shell can be considered a form of ceramic plywood! It is this tortuous crack propagation that gives rise to "graceful" fracture; the crack-bridging inherent in the fracture process is so marked that failed samples are still more-or-less intact after the fracture event!

Fracture toughness, K_{Ic}, has also been determined for the nacreous and prismatic structures. Simple

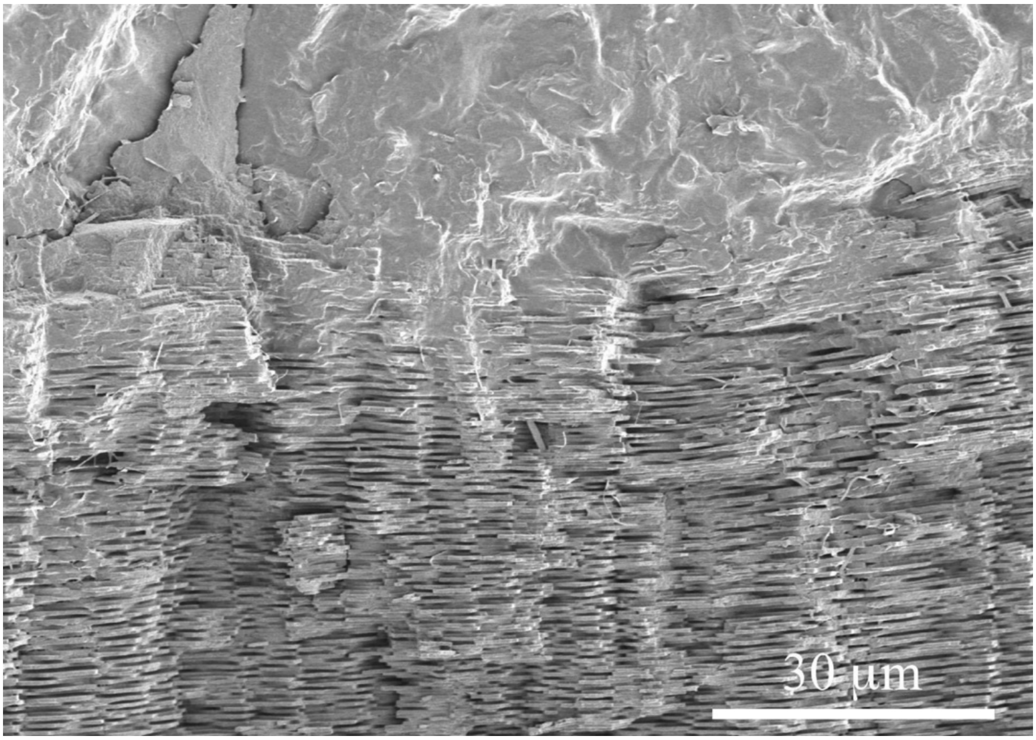


Figure 6
SEM image of the fracture surface of the red abalone shell, *Haliotis rufescens*. The upper part of the image is the calcitic prismatic portion of this shell, the lower part the aragonitic nacreous portion.

fracture-mechanics specimens were used, which may not be appropriate for such tough materials. For nacre from *Pinctada*, Jackson *et al.* (1988) reported values of 3.3 MPam$^{1/2}$ and 4.9 MPam$^{1/2}$ for two different specimen geometries, while Sarikaya and Aksay (1992) reported a value of 8 MPam$^{1/2}$ for nacre from red abalone. We have also determined K_{Ic} for nacreous and prismatic structures of the red abalone, *Haliotis rufescens*; the nacreous layer is tougher, with K_{Ic} being 2.2 ± 0.8 MPam$^{1/2}$ compared with K_{Ic} for the prismatic layer, 0.8 ± 0.4 MPam$^{1/2}$. The fracture surface of a specimen from this shell is shown in Fig. 6; the much more tortuous path taken by the crack through the nacreous layer is apparent and correlates well with the higher toughness.

See also: Marine Teeth (and Mammal Teeth)

Bibliography

Bøggild O B 1930 The shell structure of the mollusks. *K. Danske Vidensk. Selsk, Skr.* **2**, 232–325

Gregoire C 1972 Structure of molluscan shell. In: Florkin M, Scheer B T (eds.) *Chemical Zoology*. Academic Press, New York, pp. 45–103

Jackson A P, Vincent J F V, Turner R M 1988 The mechanical design of nacre. *Proc. R. Soc. London* **B234**, 415–40

Jackson A P, Vincent J F V, Turner R M 1990 Comparison of nacre with other ceramic composites. *J. Mater. Sci.* **25**, 3173–8

Kamat S 2000 Toughening mechanisms in laminated composites: a biomimetic study in mollusk shells. PhD Thesis, Case Western Reserve University

Kamat S, Su X, Ballarini R, Heuer A H 2000 Structural basis for the fracture toughness of the shell of the conch *Strombus gigas*. *Nature* **405**, 1036–40

Kuhn-Spearing L T, Kessler H, Spearing S M, Ballarini R, Heuer A H 1996 Fracture mechanisms of the *Strombus gigas* conch shell: Implication for the design of brittle laminates. *J. Mater. Sci.* **31**, 6583–94

Laraia V J, Heuer A H 1989 Novel composite microstructure and mechanical behavior of mollusk shell. *J. Am. Ceram. Soc.* **72** (11), 2177–9

Laraia V J, Heuer A H 1990 Microstructures and mechanical behavior of mollusk shells. In: Bentzen J J, Bilde-Sorensen J B, Christiansen N, Horsewell A, Ralph B (eds.) *Proc. 11th RISØ Int. Symp. Metallurgy and Materials Science, Structural Ceramics—Processing, Microstructure, and Properties*. RISØ National Laboratory, Roskilde, Denmark

Laraia V J, Aindow M, Heuer A H 1990 An electron microscopy study of the microstructure and microarchitecture of the *Strombus gigas* shell. *Mater. Res. Soc. Symp. Proc.* **174**, 117–24

Nakahara H G, Bevelander, Kakei M 1982 Electron microscopic and amino acid studies on the outer and inner shell layers of *Haliotis rufescens*. *Venus, Kyoto* **39**, 205–11

Nakahara H 1983 Calcification of gastropod nacre. In: Westbroek P, de Jong E W (eds.) *Biomineralization and Biological Metal Accumulation*. Reidel, Dordrecht, The Netherlands, pp. 225–30

Sarikaya M, Aksay I A 1992 Nacre of abalone shell: a natural multifunctional nanolaminated ceramic–polymer composite. In: Case S T (ed.) *Results and Problems in Cell Differentiation*. Springer, Berlin, pp. 1–26

Schäffer T E, Ionescu-Zanetti C, Proksch R, Fritz M, Walters D A, Almqvist N, Zaremba C M, Belcher A M, Smith B L, Stucky G D, Morse D E, Hansma P K 1997 Does abalone nacre form by heteroepitaxial nucleation or by growth through mineral bridges? *Chem. Mater.* **9**, 1731–40

Weiner S, Traub W, Lowenstam H A 1983 Organic matrix in calcified exoskeletons. In: Westbroek P, de Jong E W (eds.) *Biomineralization and Biological Metal Accumulation*. Reidel, Dordrecht, The Netherlands, pp. 205–24

A. H. Heuer, X. Su, S. Kamat and R. Ballarini
Case Western Reserve University, Cleveland, Ohio, USA

Si_3N_4 Ceramics, Structure and Properties of

Silicon nitride ceramics are composite materials, consisting of Si_3N_4 grains embedded in a matrix of amorphous or partially devitrified glass. They represent a whole class of materials with various properties, which depend on the volume fraction and composition of the grain-boundary phase as well as on the size and morphology of the Si_3N_4 grains. The microstructure is usually characterized by elongated grains surrounded by a fine-grained matrix. Since the elongated grains grow during the densification process, the resulting type of microstructure has been labeled as either self-reinforced or *in situ* toughened. Due to the self-reinforced microstructure, silicon nitride ceramics can exhibit an exceptionally high strength and toughness and make them interesting as structural components for numerous applications in engines and turbines. Another steadily growing field of application is the use-as-wear parts in ball bearings and pumps, and they are widely used as cutting tools for the machining of cast iron (Chen *et al.* 1993, Hoffmann and Petzow 1994a).

1. Crystal Structures

Silicon nitride has a density of 3180 kg m^{-3}. During heating, it decomposes into liquid silicon and nitrogen gas at 1877 °C. The basic structural units are highly covalent bonded SiN_4 tetrahedra. They can be linked together in different ways to form three different modifications. Most recently, a high-pressure modification with cubic symmetry has been discovered by Zerr *et al.* (1999). The most stable phase is denoted as β-Si_3N_4 and crystallizes in the hexagonal space group $P6_3$. The α-Si_3N_4 also has hexagonal symmetry (P31c), but it is metastable and forms only at lower temperatures.

There is some evidence that α-Si_3N_4 is indeed a defect structure with a minor replacement of nitrogen atoms by oxygen atoms (Jack 1976). The two forms cannot be transformed one to the other by a displacive transformation and a reconstructive process is necessary, usually in the presence of a liquid phase. The exact nature of the phase transformation is still not known. Both Si_3N_4 modifications can form solid solutions with Al_2O_3 (see *Si-AlON Ceramics, Structure and Properties of*).

2. Processing and Microstructural Development

In general, silicon nitride ceramics are manufactured by standard powder metallurgical routes. The powder compacts are prepared by cold pressing, slip casting, or injection molding. Two basic types of silicon nitride ceramics are known: sintered silicon nitrides (SSN), and reaction-bonded silicon nitrides (RBSN). RBSN are prepared from silicon powder compacts, which are nitrided at 1100 °C to 1450 °C in a nitrogen atmosphere, usually under atmospheric pressure. The temperature–time program of the exothermic reaction depends on the geometry of the components and the particle size of the silicon powder. The final product shows a high porosity and, therefore, a low strength, which limits the range of applications (Ziegler *et al.* 1987). SSN are prepared from silicon nitride powder compacts. Due to the highly covalent bonding of the silicon nitride, densification can only be obtained by liquid-phase sintering using metal oxides such as MgO, Al_2O_3, Y_2O_3, or rare earth as sintering additives.

Most of the sintered silicon nitride ceramics are prepared by using α-rich Si_3N_4 powders which transform during sintering into β-Si_3N_4. The β-modification forms elongated grains if there is no pronounced grain impingement. The grain-growth anisotropy with the fast growth direction parallel to the crystallographic *c*-axis leads to a needle-like grain morphology with aspect ratios of up to 10, similar to whiskers. Since homogeneous nucleation of β-Si_3N_4 did not occur during sintering, the volume fraction of the elongated grains and their aspect ratio is determined by the number and size of the β-seeds in the starting powder (Hoffmann and Petzow 1994b).

A typical microstructure of an *in situ* toughened sintered silicon nitride ceramic is shown in Fig. 1. By using pressure-assisted densification methods with pressures between 10 MPa (gas-pressure sintering) and 100 MPa (hot isostatic pressing), the sintering temperature could by raised above the decomposition temperature of silicon nitride and complete densification could be achieved even with small amounts of additives (2 vol.%). Higher sintering temperatures also promote grain growth and an increase of the volume fraction of elongated grains. The second effect of an applied pressure is the reduction of the critical flaw size in the final product, which gives in an improvement in strength and reliability (Mitomo and Uesono 1992).

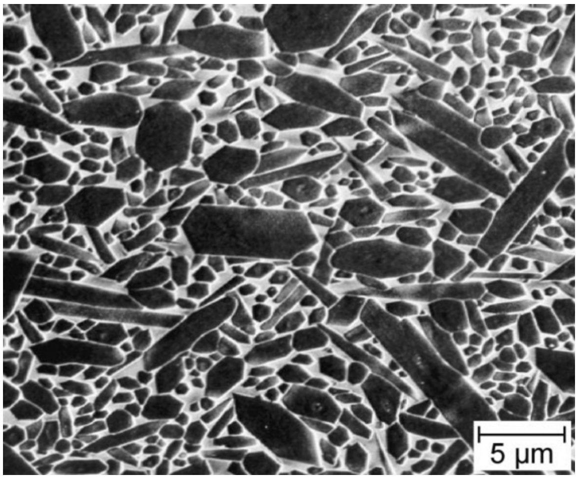


Figure 1
Scanning electron micrograph of a plasma-etched silicon nitride microstructure. The dark particles are silicon nitride grains and the bright phase is the amorphous grain boundary.

An alternative manufacturing method for silicon nitride ceramics is based on a combination of the RBSN and SSN processes. The initial powder compacts consist of silicon powder and sintering additives. The silicon powder is first nitridated and subsequently postdensified at temperatures between 1750 °C and 1850 °C. The final product is denoted as SRBSN; it has a similar microstructure to the sintered silicon nitride ceramics (SSN) and comparable properties.

One of the main features of a SSN microstructure is the complete wetting of the grain boundaries. Two-grain junctions between adjacent grains are separated by intergranular films of a constant thickness, as shown in Fig. 2. The equilibrium film thickness is determined by the additive composition and varies between 1 nm and 2 nm. The intergranular films are always amorphous, whereas pockets at multigrain junctions could be crystallized (Kleebe *et al.* 1992). High-temperature resistant silicon nitride materials mostly have crystallized multigrain junctions, while materials for applications in the temperature range below 1000 °C often reveal a completely amorphous grain-boundary phase.

The properties of these intergranular films have a strong impact on the strength and toughness at room temperature, since *in situ* reinforcement not only requires grains with high aspect ratios, but also a tailored interfacial strength which has to be weak enough so that toughening mechanisms can be activated upon failure. There seems to be a contradiction between the development of high-strength or

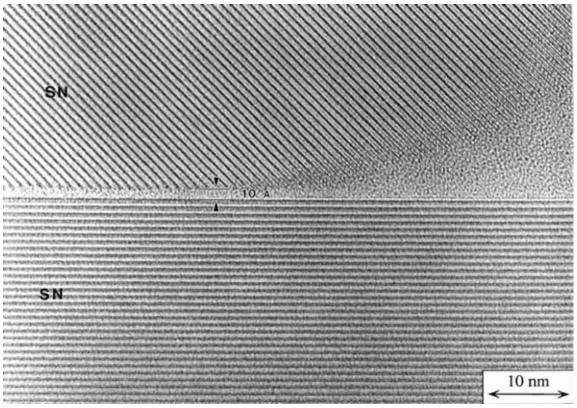


Figure 2
High-resolution transmission electron micrograph of an intergranular grain boundary film of 1 nm thickness between two parallel silicon nitride grains (left side), with a transition to an amorphous multigrain junction on the right side.

high-toughness materials and high-temperature resistant ceramics due to the fact that the intergranular film is on one hand responsible for the exceptional properties at lower temperatures, but on the other hand it limits the properties at temperatures above the softening point of the film. Indeed, high-temperature resistant silicon nitrides only reveal an intermediate strength and toughness level because of the stronger interfacial bonding between the films and the silicon nitride grains.

3. Mechanical Properties

Toughening in silicon nitride ceramics with a high volume fraction of elongated grains and a weak interface between the glass matrix and the grains is attributed to crack-wake mechanisms such as crack bridging, grain rotation, and grain pull-out, as well as crack-tip mechanisms such as debonding or crack deflection. For structural applications, the fracture conditions will typically be governed by the presence of small flaws or defects present either in the as-sintered component or generated under in-service stresses. Because toughening mechanisms develop and become increasingly effective as a small crack extends and interacts further with the microstructure, the fracture resistance increases with crack size giving rise to so-called R-curve behavior. The rate of increase of the fracture resistance (steepness of the R-curve) with the initial extension of a small crack is quite important for catastrophic failure characteristics as the fracture-origin defects are typically below 100 μm in size for these type of materials (Becher 1991, Becher *et al.* 1998).

Since fine-grained silicon nitrides exhibit a very steep R-curve with a maximum toughness up to $12\,\mathrm{MPa\,m^{1/2}}$, a high strength up to 1400 MPa with a Weibull modulus between 10 and 30 could be obtained. However, there is significant strength degradation at temperatures above 1000 °C, when the intergranular films start to soften. Materials developed for high-temperature applications have a lower mean bending strength between 600 MPa and 800 MPa, and a fracture toughness between $5\,\mathrm{MPa\,m^{1/2}}$ and $6\,\mathrm{MPa\,m^{1/2}}$ due to the stronger interfacial bonding (Hoffmann 1995). The best high-temperature-resistant silicon nitrides can be used up to 1400 °C for several hundred hours without damage by oxidation or creep. A further advantage of silicon nitride, in comparison to other structural ceramics, is the relatively high thermal shock resistance due to the low thermal expansion coefficient of $\sim 3 \times 10^{-6}\,\mathrm{K^{-1}}$. This means that high-strength silicon nitride ceramics can withstand temperature differences of more than 800 °C.

4. Conclusions

Silicon nitride ceramics have several excellent properties and clear advantages over other structural materials, but many potential applications are still not realized. A breakthrough in their use is inhibited not only by the high costs of the components for the automotive market; but also by a large scatter of the property data such as strength, fracture toughness, and reliability, plus insufficient reproducibility. This can be attributed to the relatively complex interrelationships between the size and morphology of the Si_3N_4 grains, chemical composition and distribution of the grain-boundary phase, and the resulting properties. However, the multiphase microstructure of silicon nitride ceramics also provides a great opportunity to optimize properties for specific applications according to the demands of engineers. However, meanwhile, it is well accepted that a silicon nitride ceramic for universal use does not exist.

Bibliography

Becher P F 1991 Microstructural design of toughened ceramics. *J. Am. Ceram. Soc.* **74** (12), 1050–61

Becher P F, Sun E Y, Plucknett K P, Alexander K B, Hsueh C H, Lin H T, Waters S B, Westmoreland C G, Kang E-S, Hirao K, Brito M 1998 Microstructural design of silicon nitride with improved fracture toughness, Part I: effects of grain shape and size. *J. Am. Ceram. Soc.* **81** (11), 2821–30

Chen I W, Becher P F, Mitomo M, Petzow G, Yen T-S (eds.) 1993 *Silicon Nitride Ceramics—Scientific and Technological Advances. MRS Symposium Proceedings*. Vol. 287 Materials Research Society, Pittsburgh, PA

Hoffmann M J 1995 High-temperature properties of Si_3N_4 ceramics. *MRS Bull.* **2**, 28

Hoffmann M J, Petzow G 1994a *Tailoring of Mechanical Properties of Si_3N_4 Ceramics.* NATO ASI, Series E. Kluwer, Dordrecht, The Netherlands

Hoffmann M J, Petzow G 1994b Tailored microstructures of silicon nitride ceramics. *Pure Appl. Chem.* **66** (9), 1807–17

Jack K H 1976 Review of Sialons and related nitrogen ceramics. *J. Mater. Sci.* **11**, 1135–58

Kleebe H-J, Hoffmann M J, Rühle M 1992 Influence of secondary phase chemistry on grain-boundary film thickness in silicon nitride. *Z. Metallkd.* **83** (8), 610

Mitomo M, Uesono S 1992 Microstructural development during gas-pressure sintering of a-silicon nitride. *J. Am. Ceram. Soc.* **75** (1), 103–8

Zerr A, Miehe G, Serghiou G, Schwarz M, Kroke E, Riedel R, Fuess H, Kroll P, Boehler R 1999 Synthesis of cubic silicon nitride. *Nature* **400**, 340–2

Ziegler G, Heinrich J, Wötting G 1987 Review relationships between processing, microstructure and properties of dense and reaction-bonded silicon nitride. *J. Mater. Sci.* **22**, 3041–86

M. J. Hoffmann
Universität Karlsruhe, Germany

Si-AlON Ceramics, Structure and Properties of

Silicon nitride forms two solid solutions that have the same structures as the two forms of the parent nitride, α-Si_3N_4 and β-Si_3N_4. Commonly, α-Si_3N_4 solid solution is referred to as α-SiAlON and β-Si_3N_4 solid solution as β-SiAlON, because in both solid solutions Si–N bonds are partially replaced by Al–O bonds. This is chemically and structurally feasible since the lengths of the Si–N bond (174 pm) and the Al–O bond (175 pm) are almost identical, and their valence states are identical. Additionally, α-SiAlON contains some cations stuffed in the interstitial cages available in the α-Si_3N_4 structure. (No such cages exist in β-Si_3N_4.) This allows partial replacement of silicon by aluminum despite their different valence states and the longer Al–N bond length (187 pm).

Inasmuch as Al_2O_3 is a common additive to Si_3N_4 to aid sintering, the occurrence of either α or β solid solution is always assured whenever aluminum is present. Many so-called silicon nitride ceramics are actually SiAlON ceramics. They have substantially the same properties as α- and β-Si_3N_4 ceramics and have found similar applications. In some cases, they provide a better alternative to the parent nitride. For example, α-Si_3N_4 ceramics are difficult to obtain because they tend to transform to β-Si_3N_4 during sintering. The more stable α-SiAlON, which is harder than β-Si_3N_4 and β-SiAlON, is the only engineering ceramic that has an α-Si_3N_4 structure.

Substitution of Si–N bonds by Al–O bonds is also common in other M–Si–Al–N–O systems, where M is a metal cation. Many solid solutions exist for various compounds. Examples are SiAlONs derived from Si_2N_2O (O-phase) and $Si_3Ln_2O_3N_4$ (melilite), where Ln = lanthanide. These will not be further discussed here since they are not well known as structural ceramics.

1. Structures and Compositions of SiAlON Ceramics

1.1 History

The discovery of SiAlON ceramics was linked to the use of Al_2O_3 and other oxides, such as MgO, Y_2O_3, and Ln_2O_3, as sintering additives. Additives are needed since, for practical purposes, Si_3N_4 cannot be sintered by itself. In 1971 Oyama and Kamigaito in Japan and, in 1972, Jack and Wilson in the UK first reported the formation of a solid solution that has the same structural form as β-Si_3N_4. This is the β-SiAlON solid solution. After this report, the formation of an α-Si_3N_4 solid solution with an α structure was also observed, first in systems containing aluminum and lithium. Subsequently, other metals (M = Li, Ca, Y, and Ln with atomic number > 60) were found to enter α-SiAlON solid solution as well.

1.2 β-SiAlON

The composition of β-SiAlON can be expressed as $Si_{6-z}Al_zO_zN_{8-z}$, with z in the range 0–4.2. Thus, up to 70% of Si–N bonds can be substituted by Al–O. It is noted that there is also another compound, Al_3O_3N, corresponding to the case when all Si–N bonds are replaced. The structure of Al_3O_3N, however, is that of spinel and not β-Si_3N_4. Therefore, a complete solid solution between Si_3N_4 and Al_3O_3N is not possible. In this regard, the discovery of a spinel form of Si_3N_4 at high pressure (Zerr *et al.* 1999) is interesting since it may foretell a complete (spinel) solid solution after all.

Although the bond lengths of Si–N and Al–O are close, there is a slight extension when substitution is made, and so the unit cell of β-SiAlON expands with z. This provides a convenient means to determine the existence and the composition of β-SiAlON. The empirical correlation between lattice parameters and z for β-SiAlON is

$$a\ (\text{pm}) = 760.3 + 2.97z$$

$$c\ (\text{pm}) = 290.7 + 2.55z$$

1.3 α-SiAlON

The general formula for α-SiAlON may be written as $M_xSi_{12-m-n}Al_{m+n}O_nN_{16-n}$, where m and n correspond to the proportion of Al–N and Al–O bonds, respectively. The value of x is related to m and the valence (v) of the M cation, i.e., $x = m/v$. The lower solubility limit for x lies around 0.4, and varies with M. For lithium, it is 0.25; for calcium, 0.40; for yttrium and lanthanides, 0.33. These cations are located at the interstitial sites at $(x, y, z) = (1/3, 2/3, 3/8)$ and $(2/3, 1/3, 7/8)$. Such sites are formed in the α-Si_3N_4 structure by the operation of a c-glide plane relating the upper and lower halves of its unit cell; these halves are identical to the unit cell of β-Si_3N_4. As a result, the long vacant channel parallel to the z-axis in the β-structure is broken up into large holes repeated at the unit cell intervals. (In the z-direction, the α unit cell has a dimension that is approximately twice that of the β unit cell.) Therefore, structurally, x could be as high as 2. In reality, the maximum value of x observed strongly depends on the cation incorporated, varying from 0.6 for lanthanides (neodymium and samarium) to about 1.4 for calcium and 1.5 for lithium. The upper limit of 2 is difficult to reach and retain, except perhaps at very high temperatures.

All α-SiAlON compositions can be represented on the composition plane extended between Si_3N_4, Al_3O_3N, and $3AlN{:}M_{3/v}N$. This plane is referred to as the α-SiAlON plane. As shown in Fig. 1, it also contains β-SiAlON, which coexists with α-SiAlON of low m compositions. The apex of pure Si_3N_4 corresponds to the β phase, which proves to be more stable than the α phase in normal firing of silicon nitride ceramics. Therefore, there is a discontinuity between α-Si_3N_4 at $x = 0$ and α-SiAlON at the minimum x. In the literature, this gap has been referred to as a miscibility gap. Although it might be closed at high (or low) temperature (not kinetically accessible), for practical purposes this simply means that a minimum amount of cation interstitial is required to stabilize the α structure.

The range of α solid solution depends on the interstitial cations and temperature. Knowledge of this aspect is mostly limited to the region above the $m = 2n$ line (from Si_3N_4 to $M_{2/v}O{:}3AlN$); compositions below it involve the use of less stable metal nitrides. Generally, a larger solubility range is obtained with a smaller cation, at a lower cation valence, and at a higher temperature.

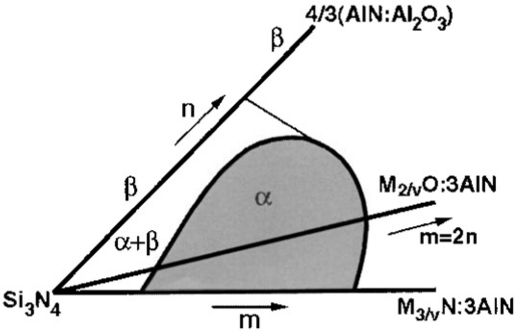


Figure 1
Composition plane containing α- and β-SiAlON.

The unit cell of α-SiAlON expands with both m and n. Note that the bond length mismatch between Al–N and Si–N is much larger than that between Al–O and Si–N. Therefore, lattice parameters are much more sensitive to m than to n. Unit cell dimensions are relatively insensitive to the type of interstitial cations. This implies that cations can all fit into interstitial cages without causing distortion. The empirical formula for lanthanide-containing α-SiAlON is

$$a\ (\mathrm{pm}) = 775.4 + 4.5m + 0.9n$$

$$c\ (\mathrm{pm}) = 562.2 + 4.0m + 0.8n$$

Because of their weak dependence on n and on interstitial cations, lattice parameters alone cannot be used to determine the composition of α-SiAlON.

2. *Thermal and Chemical Stability of SiAlON Ceramics*

2.1 Phase Stability

Both α- and β-Si_3N_4 can be successfully used as starting powders, along with other appropriate components, to form α- and β-SiAlON. Therefore, SiAlON solid solutions must have lower free energies than pure constituent phases. Generally, if α-Si_3N_4 powders rather than β-Si_3N_4 (of the same size) are used, the kinetics of SiAlON formation is faster. This is because α-Si_3N_4 is less stable, and therefore it provides a larger driving force for forming solid solutions. For β-SiAlON, a continuous solid solution extends from pure β-Si_3N_4. Thermodynamics dictates that the free energy of a solid solution must initially decrease with the amount of solute, before it rises again at larger amounts of solute. This suggests that the most stable β-SiAlON rests at some intermediate composition. Similar considerations apply to α-SiAlON. The most stable α-SiAlON should reside near the center of the single-phase region. Lastly, both α- and β-SiAlON coexist at certain compositions (see Fig. 1). Therefore, they should have comparable phase stability overall. The relative stability of α-Si_3N_4, β-Si_3N_4, and SiAlON solid solutions can thus be schematically pictured using Fig. 2.

The solubility range of β-SiAlON is essentially temperature independent. However, the solubility range of α-SiAlON is smaller at lower temperatures, especially at high m and n values. This temperature dependence is important for practical applications. When an α-SiAlON, first fired at high temperature, is re-equilibrated at a lower temperature, it is possible that some, or all, of the α-SiAlON can become unstable and dissolve, forming other equilibrium phases. For example, neodymium α-SiAlON has been found to decompose completely after being held at below

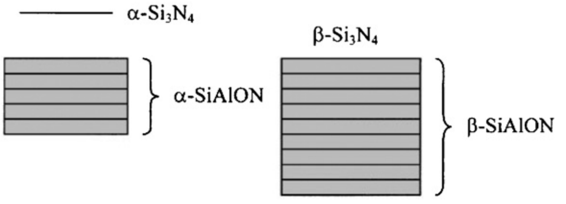


Figure 2
Schematic energy diagram for α-Si_3N_4 and β-Si_3N_4, and their solid solutions. The energy difference between α-Si_3N_4 and β-Si_3N_4 is exaggerated. Their solid solutions have comparable energies and can coexist.

1500 °C. The presence of a liquid phase facilitates the kinetics. This is usually the case when m and n values are high.

2.2 Oxidation and Chemical Resistance

Ultimately, Si_3N_4 and SiAlON ceramics are unstable against oxidation—the end products are SiO_2, Al_2O_3, and N_2. However, they are protected against catastrophic oxidation by a blocking surface oxide film. So they have lower oxidation rates than SiC. Grain boundary composition is central to the kinetics, but the stability of SiAlON solid solutions also matters. For example, oxidation resistance of α-SiAlON with different lanthanides (from neodymium to ytterbium) increases with atomic number. This is consistent with the increasing SiAlON stability in the same order. The viscosity of grain boundary glass also increases in the same order. For β-SiAlON, it is found that even 1 wt.% addition of Y_2O_3, a relatively refractory additive, increases oxidation. In two-phase α/β-SiAlON composites, independent of the sintering aids used, the most oxidation resistant ceramics are those with the highest α-SiAlON content, while β-SiAlON oxidizes the most. This trend may suggest a better stability of α-SiAlON, but the kinetics is also influenced by the presence of α-SiAlON: when additives are absorbed into α-SiAlON, the grain boundary is cleansed and hence the oxidation kinetics slowed.

Compositions of high m and n values for α-SiAlON are not desirable when good stability is called for. Simple etching would reveal that these compositions corrode easily. There is a tendency to form second phases when the amount of additives is large. Such second phases, regardless of their types, are usually not as refractory as silicon nitride and its solid solutions.

3. Fabrication of SiAlON Ceramics

3.1 Processing Methods

Both α- and β-SiAlON ceramics are fabricated by the powder processing route followed by liquid-phase sintering. The starting powders are Si_3N_4, AlN, and various oxides. In the case of β-SiAlON, only Si_3N_4, AlN, and Al_2O_3 are required but other oxide additives may be added to improve sinterability. In the case of α-SiAlON, oxides that provide interstitial cations are also required but Al_2O_3 is optional. The residual amount of oxygen in the nitride powders should be taken into account in formulation.

The native surface oxide on Si_3N_4 is the initial source of liquid. It forms a glass into which other oxides in the starting powder enter. Glass melting begins from 1200 °C to 1400 °C, depending on additives. As temperature increases, nitrides also dissolve and the amount of liquid increases. Densification then proceeds rapidly. The highest temperature used is typically between 1750 °C and 1900 °C. The process is very similar to that used for sintering silicon nitride. The main difference is that for SiAlON, the oxide component in the liquid is later absorbed into the solid solution. Therefore, the amount of the residual liquid is greatly reduced. This is important because it allows the use of a large amount of additives without leaving too much grain boundary phase that may impair properties.

Sintering of SiAlON ceramics is conducted in a nitrogen atmosphere. A nitrogen over-pressure of 10–100 atm is advantageous for it allows a higher sintering temperature to be used without causing decomposition of SiAlON. This method is referred to as gas pressure sintering to distinguish it from "pressureless" sintering in which no over-pressure is applied. Hot pressing using a die or hot isostatic pressing with or without encapsulation may also be practiced. Fabrication of dense, high-quality β-SiAlON ceramics is now routinely carried out commercially and the same for α-SiAlON ceramics may also become reality in the future.

3.2 Microstructure Control

Like silicon nitride, SiAlON ceramics may develop a microstructure of elongated grains that reinforce the structure and impart toughness against fracture. This self-reinforcement process makes silicon nitride an outstanding engineering material, with high strength and excellent resistance to thermal shock. To develop such a microstructure in β-SiAlON it is preferable to use α-Si_3N_4 powders with a low β content. The intent is to limit the population of β nuclei. With relatively few nuclei, the ones that do exist can rapidly grow, without impingement, into their characteristic shape of a hexagonal prism with a large aspect ratio. Such grains are suitable for self-reinforcement. However, if β-Si_3N_4 powders are used, then they soon build up an overcoat of β-SiAlON. The population of β-SiAlON grains is thus too high to allow many elongated grains to develop. All modern commercial β-SiAlON ceramics are produced using fine α-Si_3N_4 powders.

It was thought for a long time that the same microstructure could not be developed in α-SiAlON ceramics, for they always contained equiaxed grains and have low toughness. Subsequently, nucleation control has been successfully applied to α-SiAlON. Using fine β-Si_3N_4 starting powders with a low α content, self-reinforced α-SiAlON ceramics with elongated grains are readily obtained in most compositions (Chen and Rosenflanz 1997). Powders of α-Si_3N_4 can also be used in many cases. Most of the α-Si_3N_4 powders, having lower stability, tend to dissolve, especially if the driving force for forming α-SiAlON is not large and the formation kinetics is not fast. In such a case, a low population of α-SiAlON nuclei is again achieved and the nuclei that proceed to grow can still attain a high aspect ratio. Using these methods, self-reinforced lanthanide- and calcium-containing α-SiAlON ceramics have been developed.

A core–shell microstructure with a β-Si_3N_4 core and a β-SiAlON shell can often be seen that supports the above picture. A similar core–shell microstructure with an α-Si_3N_4 core and an α-SiAlON shell has also been observed. Detailed investigation of growth morphology suggest that α- and β-SiAlON grains can probably be grown on both α- and β-Si_3N_4 seeds, since the crystallographic structures of all the above are rather similar, at least in the *c*-direction. Nevertheless, intentional seeding is the most effective method for microstructure control. For β-SiAlON, β-Si_3N_4 seeds are adequate since they are quite stable in most β-SiAlON compositions. For α-SiAlON, α-SiAlON seeds are preferred since α-Si_3N_4 crystals from the starting powders may dissolve. The initial overgrowth usually has a composition that deviates from the overall composition because of supersaturation at the early growth stage, and because of the need to minimize misfit at the interface.

4. Thermal Mechanical Properties of SiAlON Ceramics

4.1 Intrinsic Properties

The basic thermal mechanical properties of α- and β-SiAlON ceramics are similar to the parent nitrides. Density is around 3.2 $\mathrm{Mgm^{-3}}$, Young's modulus around 310 GPa, Poisson's ratio around 0.24, and thermal expansion coefficient around 3.5% $10^{-6}\,\mathrm{K^{-1}}$. Thermal conductivity of SiAlON ceramics, however, decreases substantially as the amount of substitution increases, and is lower in α-SiAlON than in β-SiAlON. High-purity β-Si_3N_4 can achieve a thermal conductivity approaching 100 $\mathrm{Wm^{-1}K^{-1}}$, while a typical value for most α-SiAlON ceramics is only 8 $\mathrm{Wm^{-1}K^{-1}}$. The presence of metal cations at the interstitial sites obviously increases phonon scattering.

The most fundamental difference in the intrinsic mechanical properties of α- and β-SiAlON lies in their hardness. Vickers hardness, measured at 10 kg load, is around 21–23 GPa for α-SiAlON for most compositions, and decreases with z for β-SiAlON from 17 GPa at $z=0$ to about 14 GPa at $z=4$. (Composites of α/β-SiAlON ceramics have intermediate hardness between 17 GPa and 19 GPa.) This can be attributed to their different crystal structures. The stacking sequence of Si–N layers can be visualized in the ABAB sequence for the β structure, and the ABCD sequence for the α structure. The longer stacking sequence of the α structure increases the magnitude of Burgers vectors so that the slip resistance for dislocations increases. This accounts for the different macroscopic hardness.

Compared to other structural ceramics, β-SiAlON is about as hard as sapphire, whereas α-SiAlON is about as hard as SiC. Inasmuch as most applications of structural ceramics are closely related to their wear resistance, the hardness difference is important for selecting materials for engineering applications. Moreover, SiAlON ceramics retain their hardness well at high temperature. At 1100 °C, a hardness of 14 GPa can be obtained in high α content α/β composites (lower for β-SiAlON). The high hardness in the composite is due to the reduction of grain boundary glass that softens at high temperature. Since the 1970s SiAlON (mostly the β type and some α/β composites) ceramics have been used for metal cutting where local temperature can reach well above 1200 °C. Excellent hot hardness and thermal shock resistance (controlled by toughness, see below) of SiAlON ceramics are responsible for their good performance in this application.

4.2 Toughness, Strength, and Creep Resistance

Microstructure toughening is the key to high-performance silicon nitride and SiAlON ceramics. Equiaxed ceramics have low toughness, whereas ceramics that have elongated grains are *in situ* toughened. Grains of β-SiAlON are usually elongated. Tough β-SiAlON ceramics with a toughness of about 5 $\mathrm{MPam^{1/2}}$ have been commercially available for many years. They can withstand a thermal shock of about 500 °C. α-SiAlON ceramics have poor toughness because of their equiaxed microstructure. They are usually used as α/β composites in which elongated β grains provide the toughness and equiaxed α grains provide the hardness.

Advances in α- and β-SiAlON ceramics have uncovered a broad class of new *in situ* toughened materials that look promising. The steady-state toughness of modern self-reinforced α- and β-SiAlON ceramics can reach above 10 $\mathrm{MPam^{1/2}}$. As a base line, an equiaxed microstructure with a grain size of 2–3 μm or less (as in conventional α-SiAlON) provides a toughness in the range 3–4 $\mathrm{MPam^{1/2}}$. With rod-like grains of a diameter less than 1 μm and an aspect ratio of 3–4 (as in many commercial hot-pressed or gas pressure sintered β-SiAlON ceramics), toughness rises to 6–8 MPam$^{1/2}$. By deliberate seeding and alignment of grains, toughnesses up to 12 MPam$^{1/2}$ have been reported (Imamura *et al.* 2000). Indentation toughness lies at the low end (3–5 MPam$^{1/2}$) of the range mentioned above. This is because the crack length around a typical indent is short in hard ceramics; therefore, a large part of the microstructural toughening due to grain bridging is not yet developed.

In order to achieve microstructure toughening, the interface of elongated grains must debond easily. The strength of the interface is sensitive to microchemistry. High-purity β-Si_3N_4 with nothing else but SiO_2 debonds poorly and has low toughness. However, when β-Si_3N_4 grains are enveloped by a β-SiAlON layer, bonding between the grain boundary Si–Al–O–N glass and the Si_3N_4 crystal is also strengthened to impair toughness (Sun *et al.* 1998). Postsintering heat treatment can improve toughness by altering impurity segregation, causing phase precipitation, and modifying residual stresses. Such treatment, however, may generate microcracks that lower the strength.

Common self-reinforced α- and β-SiAlON ceramics have a strength of 600–900 MPa at room temperature and 200–400 MPa at 1200 °C. Quite often the strongest SiAlON ceramics at room temperature are the weakest at high temperature. This is because additives that enhance sintering the most tend to provide a ceramic of highest density with fewest flaws, hence the highest strength at low temperatures. However, such additives compromise the refractoriness of grain boundary glass, making it susceptible to softening at high temperature, hence the low strength. Among common additives to silicon nitride and SiAlON ceramics, Al_2O_3 has a much more adverse effect on creep strength than rare-earth oxides. The most creep resistant (but only moderately tough) commercial nitrogen ceramics are not SiAlON; they contain only rare-earth oxides (Y_2O_3 and Yb_2O_3), not Al_2O_3.

Composites of α/β-SiAlON ceramics are attractive because they combine β phase toughening and α phase hardening; in addition, α-SiAlON absorbs additives and reduces the amount of grain boundary phases. Hot-pressed composites made from Si_3N_4, AlN, and Y_2O_3 have very clean grain boundaries. Strengths of 1.4 GPa at room temperature and 1 GPa at 1400 °C have been reported in such composites (containing 15% α-SiAlON) (Wada and Ukyo 1991). If such outstanding high-temperature strength can be achieved in self-reinforced α-SiAlON, it may become the ultimate nitrogen ceramic for structural applications.

Bibliography

Cao G Z, Metselaar R 1991 α'-Sialon ceramics: a review. *Chem. Mater.* **3**, 242–52

Chen I W, Rosenflanz A 1997 A tough SiAlON ceramic based on α-Si_3N_4 with a whisker-like microstructure. *Nature* **389**, 701–4

Ekstrom T, Nygren M 1992 SIAlON ceramics. *J. Am. Ceram. Soc.* **75** (2), 259–76

Imamura H, Hirao K, Brito M E, Toriyama M, Kanzaki S 2000 Further improvement in mechanical properties of highly anisotropic silicon nitride ceramics. *J. Am. Ceram. Soc.* **83** (3), 495–500

Hoffmann M J, Petzow G 1994 *Tailoring of Mechanical Properties of Si_3N_4 Ceramics*, NATO ASI Series, E 276. Kluwer, Dordrecht, The Netherlands

Jack K H 1993 SiALON ceramics: retrospect and prospect. *MRS Symp. Proc.* **287**, 15–27

Sun E Y, Becher P F, Plucknett K P, Hsueh C H, Alexander K B, Waters S B, Hirao K, Brito M E 1998 Microstructural design of silicon nitride with improved fracture toughness: II. Effects of yttria and alumina additives. *J. Am. Ceram. Soc.* **81** (11), 2831–40

Sun W Y, Tien T Y, Yen T S 1991 Solubility limits of α'-SiAlON solid solutions in the system Si,Al,Y/N,O. *J. Am. Ceram. Soc.* **74** (10), 2547–50

Wada S, Ukyo Y 1991 Microstructure and properties of α/β-SiAlON composite. *Proc. 13th Japanese Congress on Materials Research.* Society of Materials Science, Kyoto, Japan, 29–33

Zerr A, Miehe G, Serghiou G, Schwarz M, Kroke E, Riedel R, Fuess H, Kroll P, Boehler R 1999 Synthesis of cubic silicon nitride. *Nature* **400**, 340–2

I.-W. Chen and R. Shuba
University of Pennsylvania, Philadelphia,
Pennsylvania, USA

Stainless Steels: Cast

Stainless steels are usually defined as alloys which contain iron and a minimum of 12%, by weight, of chromium. Stainless steels can be broadly considered to be either corrosion resistant or heat resistant. In the USA, cast grades have been assigned a designation system which differentiates between corrosion and heat resistant grades. Corrosion resistant grades are used in aqueous media where the primary concern is the ability to resist corrosion and the mechanical properties are of secondary importance. Heat resistant alloys are used at high temperatures where creep resistance is of primary importance and resistance to corrosion is a secondary but important concern.

1. ACI Designations

The designation which is most commonly used in the USA is the ACI (from Alloy Casting Institute, which is now part of the Steel Founders' Society of America) system. The ACI designation system is explained as follows.

1.1 Corrosion Resistant Alloys

The alloy designation CF8M is one of the most common stainless steel grades. The letter C indicates that the alloy is a corrosion resistant alloy. The F indicates the chromium and nickel content of the alloy; it should be pointed out that this second letter only gives an indication of the chromium and nickel contents, it does not imply a single defined level of chromium or nickel. The number 8 indicates the maximum level of carbon for the alloy; in this case it is 0.08%. The last letter indicates which other alloys may be present in the grade; in this case it indicates that molybdenum is present. The ACI designation does not always use the chemical symbol; it may, as in this case, use only the first letter of the symbol.

1.2 Heat Resistant Alloys

HK is the designation of a heat resistant alloy. The H indicates that the grade is a heat resistant alloy. The letter K indicates the chromium and nickel content of the alloy; again it gives only an indication of the chromium and nickel contents. In some cases there may also be a number included in the designation; it will appear after the first two letters, e.g., HK40. This number indicates that it is in the middle of the carbon range. The carbon range is usually 0.10%, therefore, HK40 indicates an alloy with a carbon level between 0.35% and 0.45%. Other letters may be added after the number, which may indicate the addition of other alloying elements but this is rare in the heat resistant grades and is usually reserved for proprietary grades.

A description of the attributes of all the different grades of cast stainless steels available is beyond the scope of this article. However, cast grades are available which are similar to the wrought grades (Table 1). It should be noted that the cast heat-resistant grades, because of their higher carbon content, do not have corresponding wrought grades. A consequence of their high carbon levels is that the cast grades exhibit better creep and rupture properties, a requirement of heat resistant grades.

2. Corrosion Resistant Alloys

These grades can be categorized into the following types: ferritic, martensitic, precipitation hardening, austenitic, duplex, super austenitic, and nickel-base alloys.

2.1 Ferritic (CB30, CC50)

These materials do not respond to heat treatment. Dependent upon the alloy an improvement in mechanical properties can be obtained by balancing the composition to produce more austenite or by solid solution hardening.

2.2 Martensitic (CA15, CA15M, CA40, CA6NM)

This classification of materials can be heat treated to produce high strengths, while still having good resistance to atmospheric corrosion and organic media. The heat treatment usually consists of normalizing and tempering. The absence of the need to quench these materials and in particular CA6NM from the austenitizing temperature facilitates the manufacture of more complex castings. As a consequence of their strength levels these materials have found use in structural applications. In addition, weldability of these alloys does not create any problems.

2.3 Precipitation Hardening (CB7Cu-1, CB7Cu-2)

The precipitation hardening effect is produced in these alloys by the addition of copper. The heat treatment of these alloys is critical in producing the mechanical properties that are required. These alloys are usually solution treated at 1050 °C followed by rapid cooling to produce martensite. It is not unknown for castings to be refrigerated to maximize the amount of martensite. Applications for these alloys include airframe components, centrifuge bowls, compressor impellers, food machinery, valve bodies, and disks.

2.4 Austenitic Stainless Steels

The use of the term austenitic in cast stainless steels should not be taken to mean that they are fully austenitic. This term reflects the common comparison with their wrought equivalents. The cast alloys in this group may contain up to 20% ferrite. The most common alloys produced in this group are CF3 (304L), CF8 (304), CF3M (316L), and CF8M (316); the designations in parentheses are the wrought counterparts. Ferrite content is not limited to cast materials and can be found in wrought and weld metals.

Ferrite improves the weldability of the CF grades. If the weld metal contains a ferrite level of approximately 4%, hot cracking can be avoided. Ferrite improves the resistance to stress corrosion cracking (SCC) and intergranular (IGC) attack. The effect of ferrite in corrosive media must not be taken as a general statement as each situation must be considered on its merits.

Generally, ferrite does not have a detrimental effect on toughness of the austenitic grades as it is usually surrounded by austenite, which has high toughness. Ferrite also improves the strength level; there is more about this in the section on duplex stainless steels (Sect. 2.5).

Ferrite does have a detrimental effect on mechanical properties of all product forms when used at temperatures of 315 °C and above (ASME).

Table 1
ACI alloy designations and chemical composition ranges for heat- and corrosion-resistant castings.

			Composition(a)—% (balance Fe)							
Cast alloy designation	UNS No.	Wrought alloy type (see note A)	C	Mn	Si	P	S	Cr	Ni	Other elements
CA15	J91150	410[c]	0.15	1.00	1.50	0.04	0.04	11.5–14	1	Mo 0.5[b]
CA15M	J91151		0.15	1.00	0.65	0.04	0.04	11.50–14.0	1.00	Mo 0.15–1.00
CA40	J91153	420[c]	0.20–0.40	1.00	1.50	0.04	0.04	11.5–14	1	Mo 0.5[b]
CA6NM	J91540	F6NM[f]	0.06	1.00	1.00	0.04	0.03	11.5–14.0	3.5–4.5	Mo 0.4–1.0
CA6N	J91650		0.06	0.50	1.00	0.02	0.02	10.5–12.0	6.0–8.0	
CB30	J91803	431[c]	0.30	1.00	1.50	0.04	0.04	18–21	2	
CB7Cu-1	J92180	17-4[e]	0.07	0.70	1.00	0.035	0.03	14.0–15.5	4.5–5.5	Cb 0.15–0.35, N 0.05, Cu 2.5–3.2
CB7Cu-2	J92110	15-5[e]	0.07	0.70	1.00	0.035	0.03	14.0–15.5	4.4–5.5	Cb 0.15–0.35, N 0.05, Cu 2.5–3.2
CC50	J92615	446[c]	0.50	1.00	1.50	0.04	0.04	26–30	4	
CE30	J93423		0.30	1.50	2.00	0.04	0.04	26–30	8–11	
CF3	J92500	304L[c]	0.03	1.50	2.00	0.04	0.04	17–21	8–21	
CF8	J92600	304[c]	0.08	1.50	2.00	0.004	0.04	18–21	8–11	
CF20	J92602	302[c]	0.20	1.50	2.00	0.04	0.04	18–21	8–11	
CF3M	J92800	316L[c]	0.03	1.50	1.50	0.04	0.04	17–21	9–13	Mo 2.0–3.0
CF8M	J92900	D319(316)[c]	0.08	1.50	2.00	0.04	0.04	18–21	9–12	Mo 2.0–3.0
CF8C	J92710	347[c]	0.08	1.50	2.00	0.04	0.04	18–21	9–12	Cb 8 × C min, 1.0 max
CF16F	J92701	303[c]	0.16	1.50	2.00	0.17	0.04	18–21	9–12	Mo 1.5, Se 0.20–0.35
CG12	J93001		0.12	1.50	2.00	0.04	0.04	20–23	10–13	
CG8M	J93000	317[c]	0.08	1.50	1.50	0.04	0.04	18–21	9–13	Mo 3.0–4.0
CH20	J93402	309[c]	0.20	1.50	2.00	0.04	0.04	22–26	12–15	
CK20	J94202	310[c]	0.20	2.00	2.00	0.04	0.04	23–27	19–22	
CN7M	N08007		0.07	1.50	1.50	0.04	0.04	19–22	27.5–30.5	Mo 2.0–3.0, Cu 3.0–4.0
CN7MS	J04650		0.07	1.00	2.50–3.50	0.04	0.03	18–20	22–25	Mo 2.0–3.0, Cu 1.5–2.0
CY40	N06040		0.40	1.50	3.00	0.03	0.03	14–17	Bal	Fe 11.0
CZ100	N02100		1.00	1.50	2.00	0.03	0.03		Bal	Fe 3.0, Cu 1.25
HA			0.20	0.35–0.65	1.00	0.04	0.04	8–10		Mo 0.90–1.20
HC		446[c]	0.50	1.00	2.00	0.04	0.04	26–30	4 max	Mo 0.5[b]
HD		327[c]	0.50	1.50	2.00	0.004	0.04	26–30	4–7	Mo 0.5[b]

HE			0.20–0.50	2.00	2.00	0.04	0.04	26–30	8–11	Mo 0.5[b]
HF		302B[c]	0.20–0.40	2.00	2.00	0.04	0.04	188–23	8–12	Mo 0.5[b]
HH		309[c]	0.20–0.50	2.00	2.00	0.04	0.04	24–28	11–14	Mo 0.5[b], N 0.2
HI			0.20–0.50	2.00	2.00	0.04	0.04	26–30	14–18	Mo 0.5[b]
HK		310[c]	0.20–0.60	2.00	2.00	0.04	0.04	24–28	18–22	Mo 0.5[b]
HL			0.20–0.60	2.00	2.00	0.04	0.04	28–32	18–22	Mo 0.5[b]
HN			0.20–0.50	2.00	2.00	0.04	0.04	19–23	23–27	Mo 0.5[b]
HP			0.35–0.75	2.00	2.50	0.04	0.04	24–28	33–37	Mo 0.5[b]
HP50WZ			0.45–0.55	2.00	2.00	0.04	0.04	24–28	33–37	W 4.0–6.0, Zr 0.2–1.0
HT		330[c]	0.35–0.75	2.00	2.50	0.04	0.04	15–19	33–37	Mo 0.5[b]
HU			0.35–0.75	2.00	2.50	0.04	0.04	17–21	37–41	Mo 0.5[b]
HW			0.35–0.75	2.00	2.50	0.04	0.04	10–14	58–62	Mo 0.5[b]
HX			0.35–0.75	2.00	2.50	0.04	0.04	15–19	64–68	Mo 0.90–1.25, W 0.90–1.25,
CA28MWV	J91422	422[c]	0.20–0.28	0.50–1.00	1.00	0.030	0.030	11.0–12.5	0.50–1.00	V 0.20–0.30 Mo 1.75–2.25, Cu 2.75–3.25
CD4MCu	J93370	255[e]	0.04	1.00	1.00	0.040	0.040	24.5–26.5	4.75–6.00	Mo 3.00–4.50, N 0.10–0.30
CE8MN	J93345		0.08	1.00	1.50	0.040	0.040	22.5–25.5	8.00–11.00	Mo 1.75–2.50, N 0.15–0.25
CD6MN	J93371		0.06	1.00	1.00	0.040	0.040	24.0–27.0	4.00–6.00	Mo 2.5–3.5, Cu 1.00 max,
CD3MN	J92205	2205[e]	0.03	1.50	1.00	0.04	0.020	21.0–23.5	4.5–6.5	N 0.10–0.30 Mo 4.0–5.0, N 0.10–0.30
CE3MN	J93404		0.03	1.50	1.00	0.04	0.04	24.0–26.0	6.0–8.0	N 0.08–0.18
CF10SMnN	J92972	Nitronic 60[d1]	0.10	7.00–9.00	3.50–4.50	0.060	0.030	16.0–18.0	8.0–9.0	Mo 1.50–3.00, Cb 0.10–0.30
CG6MMN	J93970	Nitronic 50[d1]	0.06	4.00–6.00	1.00	0.04	0.03	20.5–23.5	11.5–13.5	V 0.10–0.30, N 0.20–0.40 Mo 6.0–7.0, Cu 0.5–1.00,
CK3MCuN	J93254	254SMO[d2]	0.025	1.20	1.00	0.045	0.010	19.5–20.5	17.5–19.5	N 0.180–0.24 Mo 6.0–7.0, Cu 0.75 max,
CN3MN		AL6XN[d3]	0.03	2.00	1.00	0.040	0.010	20.0–22.0	23.5–25.5	N 0.18–0.26 Mo 15.0–17.5, Fe 2.0 max,
CW2M	N26455	CA[e]	0.02	1.00	0.80	0.03	0.03	15.0–17.5	Bal	Cb 1.0 max Mo 17.0–20.0, Fe 3.0 max

Table 1
(Continued)

			Composition(a)—% (balance Fe)							
Cast alloy designation	UNS No.	Wrought alloy type (see note A)	C	Mn	Si	P	S	Cr	Ni	Other elements
CW6M	N30107		0.07	1.00	1.00	0.040	0.003	17.0–20.0	Bal	Mo 8.0–10.0, Fe 5.0 max
CW6MC	N26625	625[e]	0.06	1.00	1.00	0.015	0.015	20.0–23.0	Bal	Cb 3.15–4.50 Mo 16.0–18.0, Fe 4.5–7.5, W 3.75–5.25, V 0.20–0.40
CW12MW	N30002	C[e]	0.12	1.00	1.00	0.040	0.030	15.5–17.5	Bal	Mo 12.5–14.5, Fe 2.0–6.0, W 2.5–3.5, V 0.35 max
CX2MW	N26022	C22[e]	0.02	1.00	0.80	0.025	0.025	20.0–22.5	Bal	Mo 2.0–3.5, Fe 2.0 max, Bi 3.0–5.0, Sn 3.0–5.0
CY5SnBiM	N26055		0.05	1.50	0.5	0.03	0.03	11.0–14.0	Bal	Cu 27.0–33.0, Fe 3.50 max Cu 26.0–33.0, Fe 3.50 max,
M255	N24025		0.25	1.50	3.5–4.5	0.03	0.03		Bal	Cb 1.0–3.0
M30C	N24130		0.30	1.50	1.0–2.0	0.03	0.03		Bal	Cu 27.0–33.0, Fe 3.50 max Cu 26.0–33.0, Fe 3.50 max,
M30H	N24030		0.30	1.50	2.7–3.7	0.03	0.03		Bal	Cb 0.5 max
M35-1	N24135	400[e]	0.35	1.50	1.25	0.03	0.03		Bal	

										Cu 26.0–33.0, Fe 3.50 max, Cb 0.5 max
M35-2	N04020	400^{e}	0.35	1.50	2.00	0.03	0.03		Bal	Mo 30.0–33.0, Fe 3.00 max Mo 26.0–30.0, Fe 4.0–6.0,
N7M	N30007	B2^{e}	0.07	1.00	1.00	0.040	0.030	1.0	Bal	V 0.20–0.60
N12MV	N30012	B^{e}	0.12	1.00	1.00	0.040	0.030	1.0	Bal	Mo 3.0–4.0, Cu 0.5–1.0, W 0.5–1.0, N0.20–0.30
										CD3MWCuN
J93380		0.03	1.00	1.00	0.030	0.025	24.0–26.0	6.5–8.5	Mo 2.5–3.5, Cu 1.50–3.50, Cb 0.60–1.20	
CU5MCuC		825^{e}	0.05	1.00	1.00	0.030	0.030	19.5–23.5	38.0–44.0	Mo 2.0–3.0, N 0.10–0.20
CF3MN	J92804	316LN	0.03	1.50	1.50	0.040	0.040	17.0–22.0	9.0–13.0	

---

[a] Maximum unless range is given. [b] Molybdenum not intentionally added. [c] Common description formerly used by AISI. [d] Proprietary trademark: d1 Armco, d2 Avesta Sheffield, d3 Allegheny Ludlum Corp. [e] Common name used by two or more producers; not a trademark. [f] ASTM designation.

Note A: Wrought alloy type numbers are listed only for convenience to determine corresponding wrought and cast grades. The cast alloy composition ranges are different to the wrought composition ranges; buyers should use cast alloy designations for proper identification of castings. Note B: Most of the standard grades listed are coverd by the ASTM specifications A217, A351, A447, A451, A452, A494, A608, A743, A744, A757 and A890.

2.5 Duplex Stainless Steel

This group of alloys has been the subject of the greatest alloy development efforts since the 1980s. These alloys rely on ferrite to improve their strength levels; they have nominal levels of 50% ferrite and 50% austenite. These ferrite levels give them approximately twice the strength levels of the austenitic materials, with similar corrosion resistance. All of these alloys require controlled additions of nitrogen for control of the microstructure. Nitrogen also improves the corrosion resistance by influencing the partition of elements between the ferrite and austenite grades. A popular indicator of the relative corrosion resistance of these alloys is the pitting resistance number (PREN). The PREN is calculated from the composition:

$$\text{PREN} = \%\text{Cr} + 3.3\%\text{Mo} + 16\%\text{N}$$

The use of tungsten in some of these alloys has produced variations in the PREN calculation to include this element.

Recent work on the corrosion resistance (pitting and intergranular), and mechanical properties including impact and weldability, has shown there is no difference between the statically cast, centrifugally cast and wrought product forms (ASME).

An important factor in the fabrication of these alloys is the need to use overmatching filler metals. Ferrite levels can be influenced by the cooling rate. In the case of welds which cool rapidly there may be very high levels of ferrite, which will reduce the impact properties. To overcome this cooling rate effect the filler metals usually contain approximately 5% more nickel than the base metal. Slow cooling of the welds is not a solution to this problem as detrimental intermetallic phases will form.

This group of alloys has the potential to displace the CF grades because of its strength levels.

2.6 Nickel-base Alloys

The nickel-base alloys have followed the development of wrought alloys. The welding of these materials does not present any problems which are not already experienced in the welding of wrought materials. It is recommended that the carbon levels be held at the low end of the specifications in order to provide the optimum service performance in corrosive media.

3. Heat Resistant Stainless Steels

Generally these alloys are supplied in the as-cast condition, the only exception being some of the ferritic grades.

3.1 Ferritic (HA, HC, HD)

Higher strength levels can be achieved in HA heat treatment. Carbon may be added to HC to the order of 2.50% to give hardness levels of the order of 600 BHN (ASME). These alloys are generally used where high strength is not the primary requirement. They are used in applications where gases with high sulfur are present; in these applications the presence of large amounts of nickel would be detrimental.

3.2 Austenitic (HE, HF, HH, HI, HK, HL)

Of these alloys only HH has an ASTM specification which references the presence of ferrite. The level of ferrite is controlled by the composition of the alloy. The HH grade is most commonly supplied in the fully austenitic condition. It is common to observe sigma in all of these alloys when they have been in service.

The HK grade has been used extensively in chemical and petrochemical applications but is now largely displaced from these applications by the HP variants. The HP grade has seen the greatest development since the mid-1970s. It is rarely if ever produced in the unmodified form. The developments have been aimed at increasing rupture life and have consisted initially of the use of elements such as niobium to produce more stable carbides which facilitate longer life or the capability of working at higher stresses. The development of this alloy has been taken further by the addition of microalloying elements, primarily consisting of titanium. These have extended rupture life of 30% over the niobium modified grades.The applications for HP grades are chemical and petrochemical furnaces such as reformers and ethylene. In the case of ethylene the silicon levels may be increased to a range of 1.25–1.75% to resist carburization.

3.3 Heat Treatment Applications

The grades HT, HX, HU, and HW are used in applications where thermal cycling is imposed. This thermal cycling may be combined with operation in carburizing atmospheres. Such applications would include the heat treatment with oil quenching, and/or the carburization of products. Consequently these alloys have the ability to resist thermal fatigue and carburization. Again silicon levels may be increased to increase the resistance to carburization.

3.4 Heat Resistant Alloy Welding

The welding of all of the heat resistant alloys presents little difficulty. Welding is normally carried out without preheat or postweld heat treatment. Heating as part of the welding cycle is not recommended as it will adversely affect service life.

See also: Stainless Steels: Duplex

Bibliography

ASME ASME Boiler and Pressure Vessel Code. Section II, Table 6–360

Lundin C D, Wen S, Ruprecht W J, Monroe R W, Blair M 2000 Corrosion, toughness, weldability and metallurgical evaluation of cast duplex stainless steels. In: *Duplex America 2000 Conf.*, pp. 449–60

ASTM Specification A532-93a. In: *ASTM Book of Standards*, Volume 01.02

M. Blair
SFSA, Barrington, Illinois, USA

Stainless Steels: Duplex

As their name implies, duplex stainless steels are stainless steels containing two primary phases, f.c.c austenite and b.c.c. ferrite. These alloys were first observed in the 1920s as an outgrowth of studies on austenitic stainless steels. Their balance of good corrosion and mechanical properties has made them an important part of the stainless steel family of alloys.

1. Development of Microstructure

The development of a duplex microstructure is a natural outgrowth of the metallurgy of the Fe–Cr–Ni system, which is the basis of all stainless steels. This is illustrated in Fig. 1, which shows the 70 wt.% Fe isopleth for the Fe–Cr–Ni system, which is approximately the amount of iron present in most stainless steels. The austenite is referred to as the γ phase and the ferrite as δ. (In wrought ferritic stainless steels, the ferrite is typically referred to as α, and this nomenclature is also often used in the duplex stainless steel literature.) The more chromium in the alloy, the more ferrite will be present, and the more nickel, the more austenite will be present. Figure 2 shows a typical microstructure for a duplex stainless steel. Here the ferrite is etched darker than the austenite and is denoted as α.

Most duplex stainless steels initially solidify as δ, with the γ phase developing on cooling or during working in the $\delta + \gamma$ phase field. The amount and morphology of austenite and ferrite depends upon the exact composition of the alloy, the cooling rate after solidification, the annealing temperature, and the processing conditions. Alloying elements in addition to iron, chromium, and nickel also play a role in determining the amount of ferrite that can be developed. The exact effectiveness of any element in

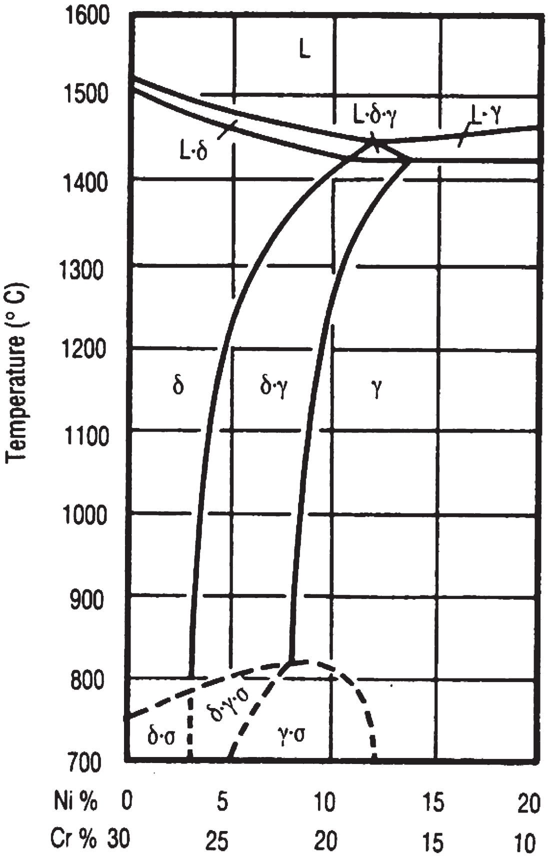


Figure 1
Pseudobinary diagram showing the 70% isoplethal section of the Fe–Cr–Ni phase diagram (reproduced by permission of EDP Sciences from 'Stainless Steels,' 1993, pp. 613–59).

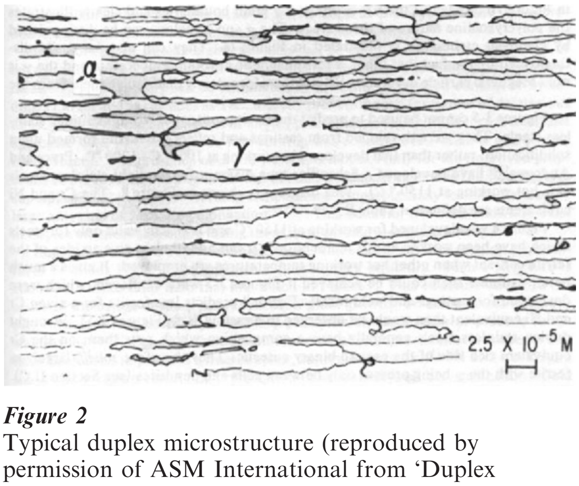


Figure 2
Typical duplex microstructure (reproduced by permission of ASM International from 'Duplex Stainless Steels,' 1983, pp. 693–756).

stabilizing ferrite or austenite varies from study to study, depending upon the exact alloy composition, processing, and heat treatment. Molybdenum, silicon, niobium, aluminum, and titanium stabilize ferrite, and manganese, copper, carbon, and nitrogen stabilize austenite. Carbon and nitrogen are particularly potent austenite stabilizers, being 15–35 times more effective than nickel (on a wt.% basis). The weighing factors for the other elements, for austenite or ferrite stabilization, are about 0.25–3.0 (also on a wt.% basis relative to nickel or chromium).

Duplex stainless steels typically contain 20–70 vol.% ferrite. They can exist in either wrought or cast forms. In addition, castings which are nominally considered as being austenitic often contain 5–40 vol.% ferrite, and are thus strictly speaking duplex stainless steels, although they are generally not referred to as such. Austenitic stainless steel weld metal is also almost always duplex, typically containing 5–20 vol.% ferrite. In the case of castings and weld metal, the ferrite is present primarily to prevent hot cracking during solidification. In the case of those alloys that are termed duplex, the ferrite is present because of its influence on corrosion and mechanical properties.

2. *Development of Additional Phases*

Duplex stainless steels can contain many phases in addition to austenite and ferrite. The development of these phases, and the way various alloying elements influence this development, is illustrated in Fig. 3. For the most part, the phases shown in Fig. 3 embrittle the alloy without strengthening it and should therefore be avoided. This is particularly true for σ and χ phases. The exception is the α' phase, which is a chromium-rich b.c.c. phase that strengthens the ferrite as well as reduces its ductility. As such, duplex stainless steels are sometimes deliberately heat treated to produce some α'. This phase can develop by either nucleation and growth or by spinodal decomposition, depending upon the composition of the alloy. The spinodal decomposition into chromium- and iron-rich regions also strengthens the ferrite.

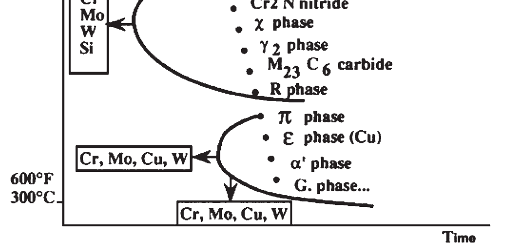


Figure 3
Schematic representation of the possible precipitates in duplex stainless steels (reproduced by permission of EDP Sciences from 'Duplex Stainless Steels' 91,' 1991, pp. 3–48).

Many of the phases illustrated in Fig. 3 develop in the ferrite or at the ferrite/austenite boundaries, so the amount and morphology of the ferrite determine the degree to which the alloy properties are degraded. Figure 3 shows the temperature ranges at which these phases form, and this sets limits on the temperature ranges at which these alloys can be fabricated or used. Rapid cooling from high-temperature processing or annealing is also required to prevent undesirable precipitates from developing.

3. *Alloy Properties and Uses*

The chief reasons for using duplex stainless steels are their good resistance to oxidation, corrosion, and stress corrosion—all this while maintaining superior mechanical properties. These alloys generally exhibit the same, or lower, general corrosion rates as austenitic stainless steels in dilute sulfuric acid. This is also true in dilute hydrochloric acid, and in the caustic solutions encountered in the pulp and paper industry. Some of these alloys are also applicable for exposure to organic acids. Many of these alloys possess superior pitting and crevice corrosion resistance, compared with austenitic stainless steels. Duplex stainless steels possess superior resistance to both transgranular and intergranular stress corrosion cracking. The exact degree to which any of these alloys is resistant to any form of corrosion or stress corrosion depends upon its composition, microstructure, and the exact nature of the solution and solution temperature to which it is exposed.

Duplex stainless steels also possess excellent mechanical properties, and are consequently sometimes utilized instead of austenitic or ferritic stainless steels. Duplex steels generally possess higher yield and ultimate tensile strengths than most austenitic or ferritic stainless steels. The degree to which this is so depends not only on the alloy composition, but also on the way that it is processed. This improved strength is generally achieved without compromising the toughness of the alloy, providing that the alloy does not contain any of the deleterious phases shown in Fig. 3. The toughness of ferrite is decreased with decreasing temperatures but, being present only as a constituent of a two-phase alloy, this does not produce as sharp a ductile-to-brittle transition as in ferritic stainless steels. The higher the ferrite content, the sharper the ductile-to-brittle transition.

Ferrite has a smaller expansion coefficient and larger thermal conductivity than austenite, so the more ferrite that is present, the lower will be the coefficient of expansion and the greater the thermal conductivity of the duplex stainless steel. Duplex stainless steels are generally very weldable; in fact, austenitic stainless steels utilize duplex weld metal.

See also: Stainless Steels: Cast

Bibliography

Charles J, Bernhardsson S (eds.) 1991 *Duplex Stainless Steels '91*. Les Editions de Physique, Les Ulis, France

Charles J 1991 The duplex stainless steels: materials to meet your needs. In: Charles J, Bernhardsson S (eds.) *Duplex Stainless Steels' 91*. Les Editions de Physique, Les Ulis, France, pp. 3–48

Desestret A, Charles J 1993 The duplex stainless steels. In: Lacombe P, Baroux B, Beranger G (eds.) *Stainless Steels*. Les Editions de Physique, Les Ulis, France, pp. 613–59

Duplex Stainless Steels '94 1994 TWI, Glasgow, UK

Duplex Stainless Steels '86 1986 Nederlands Instituut voor Lastechniek, The Hague

Gunn R N 1997 *Duplex Stainless Steels*. Abington Publishing, Cambridge, UK

Lula R A (ed.) 1983 *Duplex Stainless Steels*. American Society for Metals, Metals Park, OH

Solomon H D, Devine T M Jr. 1983 Duplex stainless steels—a tale of two phases. In: Lula R A (ed.) *Duplex Stainless Steels*. American Society for Metals, Metals Park, OH, pp. 693–756

H. D. Solomon
General Electric R & D Center, Schenectady,
New York, USA

Steels: Classifications

The beginning point for the classification of steels and iron is the iron–carbon phase diagram (see *Ferrous Alloys: Overview*). All Fe–C alloys containing less than about 2.06 wt.% carbon would pass through the austenite field when cooled slowly from the liquid to room temperature. All binary Fe–C alloys containing <2.06 wt.% carbon are classed as steels. All binary Fe–C alloys containing >2.06 wt.% carbon are termed *cast irons*. These distinctions are roughly maintained even when the alloys contain large amounts of alloying elements.

Cast irons typically have carbon levels of about 2–4 wt.% carbon. There are many types of cast irons. Most are the so-called graphitic cast irons but there are other high-alloy cast irons designed for wear resistance, corrosion resistance, and heat resistance. The graphitic cast irons (see *Cast Irons*) are alloys with the above approximate ranges of carbon content but which also contain 0.5–3.0 wt.% silicon. The purpose of the silicon is to promote the formation of graphite rather than iron carbide (cementite) on solidification or during heat treatment. Depending on the exact composition, cooling rate on solidification, and subsequent heat treatment these materials consist of particles of either cementite or graphite embedded in a steel matrix which can be entirely ferritic, ferritic–pearlitic, martensitic, or austenitic. In white cast irons the material consists of cementite in a steel matrix. The other types of cast irons consist of particles of graphite in a steel matrix. There are four different types of graphitic cast irons, gray, ductile, malleable, and compacted and these grades differ primarily in the shape of the graphite particles. In the four types the graphite shapes are plates, spheres, popcorn shaped, and rods, respectively. The strengths of each of these types of cast irons can be varied by changing the nature of the steel matrix.

For the purpose of this article steels will be regarded as all ferrous-based alloys with carbon levels in the range of 0–2 wt.%. The approach taken here is similar to that adopted by Leslie where almost all of the steels described here are discussed in greater detail (Leslie 1981).

The most widely produced steel product is sheet (see *Sheet Steel: Low Carbon*). These materials are ferrite with low carbon contents and are often used in applications requiring the sheet to be formed into complex shapes, such as auto bodies. High formability is achieved in these materials by reducing the interstitials, carbon and nitrogen, to very low levels and by developing appropriate textures through thermomechanical processing and control of inclusion characteristics. Steel sheet can be coated in a variety of ways to provide corrosion protection and for decorative purposes. These coatings reduce formability.

By adding carbon pearlite can be introduced (see *Pearlite*) into steels slowly cooled from the austenitizing temperature. When steels having less than the eutectoid carbon content are slowly cooled from the austenitizing temperature pro-eutectoid ferrite is formed as the steel is cooled to the eutectoid temperature. Once the steel is cooled below the eutectoid temperature the remaining austenite, now of the eutectoid carbon content, transforms to pearlite, which consists of alternating layers of cementite and ferrite. The amount of pearlite in the structure increases with increasing carbon content. The strength of the steel increases with the amount of pearlite and the strength of pearlite can be increased by decreasing the spacing between the alternating sheets of ferrite and cementite. Low carbon sheet steels which contain no pearlite are one extreme of the ferritic–pearlitic steel spectrum and steels of the eutectoid carbon content which are entirely pearlite when slowly cooled represent the other extreme of the ferritic–pearlitic steel spectrum. Steels of the eutectoid carbon content are commonly used for railroad tracks and wheels.

When an Fe–C alloy is cooled (quenched) sufficiently rapidly from the austenitizing temperature the austenite will not transform to a combination of pro-eutectoid ferrite and pearlite, or to bainite (see *Bainite*). Instead the alloy will transform to martensite, a body-centered-tetragonal phase, in which the carbon is in solid solution (see *Martensite*). The carbon in solid solution in the martensite is much more effective in strengthening martensite than is the pearlite in strengthening ferritic–pearlitic, steels. For this reason martensitic steels are used to achieve high strength levels. However, to achieve a martensitic structure one must cool the material sufficiently rapidly to avoid the formation of pro-eutectoid ferrite, pearlite, and bainite. Alloying additions can be used to delay the start of the decomposition of austenite to pro-eutectoid ferrite, pearlite, and bainite, and such alloying is said to increase the hardenability of the steel. When the hardenability is increased one can use slower quench rates to achieve a martensitic structure and achieve a martensitic structure in pieces of larger section size. Low alloy steels which were developed to be used primarily as martensitic steels are referred to as *heat-treated steels*. A primary concern in these steels is their hardenability. These steels are never used in their asquenched condition but are tempered. Tempering is the heating of a martensitic steel at a temperature below that at which austenite forms. Tempering results in softening of the steel due to precipitation of the carbon as cementite and the coarsening of the cementite particles. Steels used for bearings are typically low alloy quenched and tempered steels. Krauss (1980) discusses the effects of composition and heat treatment on the microstructures and mechanical properties of low alloy steels.

Ferritic–pearlitic steels have very low strength when they are purely ferritic and their toughnesses drop rapidly as the carbon content and the amount of pearlite are increased. While the martensitic quenched and tempered low alloy steels have much better toughnesses at higher strength levels than the ferritic–pearlitic steels, the quenched and tempered martensitic structure is difficult to achieve in many product forms. Costs are also associated with the heat treatment involved in obtaining the martensitic quenched and tempered microstructure. High-strength low-alloy (HSLA) steels are a class of steels developed to achieve properties superior to those of the ferritic–pearlitic steel and comparable to those of low alloy quenched and tempered martensitic steels. HSLA steels are designed to achieve their desired mechanical properties by the development of microstructures through controlled thermomechanical processing (TMP) and the steel is produced in its final form by a continuous hot deformation process, rolling, or forging, which comprise the TMP. Most HSLA steels are microalloyed with small additions (0.1 wt.%) of niobium, vanadium, and titanium which control microstructural evolution during TMP through the formation of carbonitride precipitates.

The above classes of steel have a low alloy content. Most other steels have higher alloy content and these steels were developed to achieve certain properties, often for particular applications.

The most widely used of the alloy steels are the stainless steels. These are steels with high chromium content, a minimum of about 11 wt.%. The chromium forms an oxide on the surface of the steel which is adherent and slows further oxidation or corrosion of the alloy. There are many types of stainless steels. They can be roughly distinguished on the basis of microstructure. The austenitic stainless steels are low carbon stainless steels which are austenitic at room temperature and they normally contain about 18 wt.% chromium and about 8 wt.% nickel and/or manganese to stabilize the austenite. The austenitic stainless steels are used in applications requiring corrosion resistance and at elevated temperatures when both oxidation and creep resistance are required. The martensitic stainless steels are quenched and tempered steels and there are two classes of this type of stainless steel. The first are the medium carbon steels in which strength is achieved by the precipitation of alloy carbides. These steels were developed primarily for elevated temperature service in which oxidation is a concern. The duplex stainless steels (see *Stainless Steels: Duplex*) consist of a mixture of ferrite and austenite. The ferritic stainless steels are stainless steels which are entirely ferritic. The duplex and the ferritic stainless steels each have applications in which they have better corrosion resistance or better stress corrosion cracking resistance than the martensitic or austenitic stainless steels. In addition to the wrought forms of stainless steel there are stainless steels especially made for casting. The cast stainless steels (see *Stainless Steels: Cast*) are of two types. One type is designed for corrosion resistance and the second type, the so-called *heat resistant grades*, are designed for elevated temperature service.

A wide variety of steels have been developed primarily for use at elevated temperatures. Typical tempered martensitic steels used for high temperature service have compositions (in wt.%) such as 0.15C–1Cr–0.5Mo, 0.20C–1.25Cr–0.5Mo, 0.15C–2.25Cr–1Mo, 0.15C–5Cr–0.5Mo, 0.1C–9Cr–1Mo, and 0.15C–12Cr–Mo(orW). The creep resistance of these steels increases with increasing alloy content. These alloys find applications as tubes and pipes in power plants and petrochemical plants. There has been considerable activity over the past few years in developing new 9–12 chromium steels for power plant applications both in Europe and Japan, with the primary goal to achieve improved creep resistance. While various austenitic steels have superior creep resistance to the martensitic steels the martensitic steels have lower coefficients of thermal expansion and higher heat conductivities than do the austenitic steels and this results in the martensitic steels having better fatigue resistance than the austenitics at elevated temperatures. Viswanathan (1989) discusses the high temperature properties of steels.

Steels developed primarily to achieve high toughness are of two types. The first are alloys developed to have very low ductile-to-brittle transition temperatures so that they can be used at very low temperatures. The prototype martensitic steel used in cryogenic applications are the so called *9-nickel steels*. A variety of methods can be used to lower the ductile-to-brittle transition temperature, including refinement of grain size, lowering the yield strength, adding nickel and introducing austenite of the appropriate morphology, and mechanical stability. The second are steels designed to achieve high toughness at very high strength levels. Steels of this type are the precipitation strengthened maraging steels and low to medium carbon secondary hardening steels which have been developed according to methods used by Speich and co-workers in the development of the steel HY180 (Speich *et al.* 1973).

A variety of steels have been developed to achieve various types of wear resistance. These include the Hadfield steels, tool and die steels (see *Tool and Die Steels*) and high speed steels.

The Hadfield steels are medium to high carbon steels containing substantial amounts of manganese. These alloys are austenitic at room temperature. These alloys have excellent wear resistance because the austenite transforms to a high carbon martensite when the material is subjected to strain. Thus wear results in a very hard surface layer which is resistant to further wear.

Tool Steels (Roberts and Cory 1980) covers three classes of steel: tool steels, die steels (both cold-work and hot-work), and the high speed steels. Tool steels are traditionally divided into three categories. Carbon tool steels, low-alloy tool steels, and the special purpose tool steels, of which there are several types. There are four types of cold-work die steels, oil hardening cold-work die steels, air hardening cold-work die steels, high carbon–high chromium cold-work die steels, and wear resistant cold-work die steels. There are five classes of hot-work steels, 3–4 wt.% chromium hot-work die steels, chromium–molybdenum hot-work die steels, chromium–tungsten hot-work die steels, tungsten hot-work die steels, and molybdenum hot-work die steels. Many of the steels contain large amounts of strong carbide-forming elements for two reasons. The first is to provide for a dispersion of larger carbides which are in the steel after austenitizing and which improve wear resistance. In addition, the strong carbide forming elements which are in solid solution after austenitizing can combine with carbon during tempering and this precipitation of alloy carbides (secondary hardening) provides strength at room and elevated temperatures.

High-speed steels are used primarily for the manufacture of cutting tools. While there are a number of classes of high speed steels they have in common high carbon contents (ranging from 0.8 wt.% to >3 wt.%), chromium levels of about 4 wt.%, massive amounts of strong carbide forming elements (vanadium, tungsten, and molybdenum) to provide large carbides which are present after austenitizing and which improve wear behavior and to provide for secondary hardening on tempering. Many high-speed steels also contain large amounts of cobalt, which enhances the secondary hardening response and the hot hardness of the secondary hardened microstructure.

Steels developed primarily for particular electrical, magnetic, and physical properties include the silicon steels, magnetic steels and steels having low coefficients of thermal expansion.

Silicon steels are ferritic alloys of iron and silicon that have magnetic properties which make them useful in motors and transformers. The silicon additions improve magnetic softness and increase the electrical resistivity. They also have the undesirable effects of decreasing the Curie temperature, reducing the saturation magnetization, and of embrittling the alloy when the silicon additions exceed about 2 wt.%. The embrittling effects of silicon make it difficult to produce silicon steels with more than about 3 wt.% silicon. The silicon steels are produced in two forms, highly textured grain-oriented alloys and alloys in which the grains are not oriented. Grain orientation is carried out to align the magnetic easy axis.

There are a large number of iron-based permanent magnet alloys. In addition, there are a number of iron-based soft magnetic materials in addition to the silicon steels. These alloys include Fe–Co alloys, permalloy and iron-based amorphous and nanocrystalline soft magnetic materials.

Iron-based alloys having near-zero or even negative expansions on heating in the vicinity of room temperature were first discovered in the late 1900s by the Swiss physicist Charles-Edouard Guillaume. The family of alloys which possess this characteristic are known as "Invar" for length INVARiant. The first Invar alloy had a composition of iron with 36 wt.% nickel. This alloy has a coefficient of thermal expansion varying from slightly negative to less than 10% of that of a normal metal from near absolute zero to near the Curie temperature. In Invar there is a volume contraction as the alloy transforms from ferromagnetic to diamagnetic which almost exactly counterbalances the thermal expansion. By varying the nickel content and alloying with additional elements Invar alloys have been developed which match the coefficients of thermal expansion of certain glasses and of silicon.

Bibliography

Krauss G 1980 *Principles of Heat Treatment of Steel*. ASM International, Metals Park, OH

Leslie W C 1981 *The Physical Metallurgy of Steels*. McGraw-Hill, New York

Roberts G A, Cary R A 1980 *Tool Steels*. ASM International, Metals Park, OH

Speich G R, Dabkowski D S, Porter L F 1973 Strength and toughness of Fe–10Ni alloys containing C, Cr, Mo, and Co. *Metall. Trans.* **4**, 303–15

Viswanathan R 1989 *Damage Mechanisms and Life Assessment of High-Temperature Components*. ASM International, Metals Park, OH

W. M. Garrison Jr
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Structure of Polymer Glasses: Short-range Order

Polymer glasses may be viewed at the simplest level as frozen macromolecular liquids, in which the structure at both long and short length scales parallels that found in the molten state. Thus, a polymer glass is characterized by the absence of long-range order, for example there is no orientational order as in the liquid crystal state, or translational order as in the crystalline state. There will be order along the polymer chain as is expected on the basis of the chemical connectivity, this will be most specific for a stereoregular homopolymer, and some short-range liquid-like order due to the requirement of nonoverlapping chain segments that occurs in all liquids, polymeric or otherwise.

Polymers exhibit this disordered solid state either as a consequence of a thermal history, which includes cooling from the melt state at a rate which inhibits crystallization, or through the presence of disorder along the polymer chain such as present in random copolymers or atactic systems. Thus, some materials such as polystyrene and polymethylmethacrylate are ubiquitously present in the glassy state, while others such as glassy polyethylene can only be prepared through special treatment and can never be observed at room temperature.

The study of short-range order of glassy polymers is concerned with structural arrangements in a polymer glass on length scales, which directly involve the atomistic chemical detail of the polymer, and which extend to include the chain segment and its interaction with other chain segments. A chain segment is a conceptual object that may be usefully related to the length over which the polymer chain has a persistent conformation. The size of chain segment may vary from a few skeletal bonds (~5 Å) for polyethylene to 20–30 Å for polycarbonate. As a consequence, studies of short-range order or local structure are concerned with length scales from 1 Å to 50–100 Å.

The lack of long-range order means that most glassy polymers are optically transparent. A more critical test is the nature of the so-called wide-angle x-ray scattering pattern as shown in Fig. 1; the absence of sharp Bragg reflections and the presence of broad diffuse features is the fingerprint of a glassy polymer. X-ray and neutron scattering together with nuclear magnetic resonance (NMR) procedures are the most commonly employed experimental techniques in the study of short-range order in polymer glasses, although both are often tightly coupled to

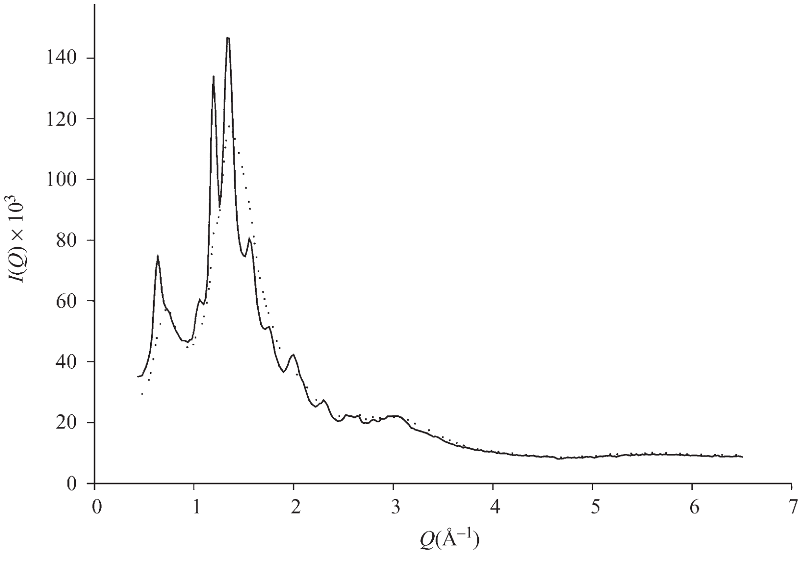


Figure 1
The wide-angle x-ray scattering patterns for a crystalline sample of isotactic polystyrene (solid line) and glassy isotactic polystyrene (broken line) formed by rapidly cooling from the melt at 160 °C.

atomistic level computational models. The literature abounds with results, in fields as diverse as diffusion, guest–host systems for nonlinear optics, mechanical properties and blends, which the authors claim can only be understood by the presence of ordered regions within these apparently amorphous materials. Clearly the local structure including spatial and orientational correlations must reflect the shape and size of the chain segments as in any molecular fluid. The issue in the study of the short-range order of polymer glasses is whether the system is ordered over and above the natural minimum of order imposed by the detail of the chain segment.

1. *Experimental Techniques*

Both NMR techniques and wide-angle x-ray and neutron scattering procedures are sensitive to variation in local structure. NMR spectroscopy provides a powerful method to examine the structure and dynamics of polymers; its main advantage is the unprecedented chemical selectivity available. Although polymer solutions can be studied with high resolution, the angular-dependent anisotropic interactions present in the solid state drastically reduce the resolution available. Although various methodologies have been employed, it has been the advent of multidimensional NMR that has driven the development of techniques for the study of the solid state (Schmidt-Rohr and Spiess 1994). By employing particular sequences of pulses the broad features present in standard solid-state NMR spectra can be probed in a second frequency domain providing great separation of features. These techniques have been employed in a variety of studies in both glassy polymers and polymer blends in which the information obtained is specific to particular bonds or groups and very localized in its extent. In contrast, scattering techniques provide information over a range of length scales but which requires models or advanced techniques such as isotopic substitution and neutron scattering to enable features to be related to particular interactions.

The measurement of x-ray and neutron scattered intensities over a broad Q ($Q = 4\pi\sin\theta/\lambda$, where 2θ is the scattering angle and λ the incident wavelength, $Q_{\min} \sim 0.2\,\text{Å}^{-1}$) provides information about bond lengths to chain segment interactions. Neutron scattering offers the possibility of high Q values, which increases the real space resolution and variations in scattering lengths through isotopic substitution. X-ray scattering instrumentation is more readily available, often with a much higher resolution in terms of ΔQ and with a contrast that arises from differences in electron density and hence is strongly related to the atomic number. For example, the x-ray scattering pattern of polyvinyl chloride (PVC) is dominated by the distribution of the chlorine atoms, while the neutron scattering pattern provides information more or less evenly weighted between the carbon, hydrogen, and chlorine atoms.

Analysis of the scattering patterns may either take place in reciprocal (experimental) space or in real space (Mitchell 1989). Real space functions representing the averaged correlations between atoms are obtained by Fourier transformation of the scattered intensities. Although this appears to be an attractive option since it allows strongly correlated distances to be read off, such functions often only serve to emphasize the well-defined covalently bonded distances, which of course are usually well known. Analysis of the reciprocal or scattering space data is usually simplified in that data for $Q > 3\,\text{Å}^{-1}$ is largely intrachain in origin and provides information on the conformation and chemically defined structures.

Data at lower scattering vectors arises principally from interchain interactions. The precise relationship between the location of peak maxima (Q_0) and the corresponding real space distances (d) is rather complex, however we can take as a working relationship that $d = k2\pi/Q_0$ where k is constant in the range 1–1.2 (Mitchell 1989). Thus, in Fig. 1 the scattering curve shows an intense broad peak at $Q \sim 1.4\,\text{Å}^{-1}$ which may be attributed to the correlation between chain segments whose "size" is $\sim 5\,\text{Å}$. The width of this peak at half-height is $\sim 0.5\,\text{Å}^{-1}$, which, using the expression $\Delta Q = 2\pi/L$, gives L the correlation length of $\sim 12\,\text{Å}$. These values are typical of the length scales involved in polymer glasses.

2. *Computational Approaches*

The increasing power of computers makes the generation of realistic atomistic level models of glassy polymers a genuine possibility (Monnerie and Suter 1994). However, for most glass-forming systems, the size of models necessary to accommodate both the scale of the chain segment and the inherent chemical inhomogeneity of the polymer chain extend beyond current capabilities. In many cases, the validation of published models, through, for example, comparison of a calculated x-ray scattering pattern with that obtained experimentally, is at best cursory. Approaches in which a coarse-grained system is decorated with atomistic detail provide one promising route forward (Kotelyanskii *et al.* 1996). A common approach is the so-called amorphous growth in which units are successively added to a growing chain within periodic boundary conditions. If the chain encounters severe steric overlap, the system backtracks until the next unit can be added at a relatively low energy cost. The difficulty here is that the larger chain configuration is often distorted from experimental values (Clark and Brown 1994). For glassy models, subsequent energy minimization or molecular dynamics only yields local minima, which may be a long way from a realistic structure.

3. Basic Observations on the Short-range Order in Polymer Glasses

Figure 2 shows the broad *Q* neutron scattering for 1,4 polybutadiene as a function of temperature. Over the complete temperature range the high *Q* scattering (intrachain correlations) is largely invariant, while the intense but diffuse peak at $Q \sim 1.4\,\text{Å}^{-1}$ (interchain correlations) show systematic variation with temperature. The position of this intense peak reflects the variation in density of the material; there is little movement in the glassy state, marked variation in the liquid state with an abrupt change at the glass transition (Fig. 3(a)). In a similar vein, the width of the intense peak decreases with temperature reflecting the thermally stimulated dynamics of the chain segments (Fig. 3(b)). These observations reflect the basic model of glassy polymer as a frozen liquid and hence we can transfer concepts such as the random coil model from the melt.

Polymer glasses are usually homogenous as indicated by their optical clarity and by small-angle x-ray scattering measurements, which show that the only significant density variations are those arising from thermal fluctuations (Wendorff 1985). The literature abounds with other reports that present conflicting evidence for ordered structures, the Faraday Discussion in 1979 and Gabrys' review (1994) provide good starting points; the current consensus is that most of this historical evidence is unreliable. The study of the static structures of polymer glasses overlaps with analysis of polymer dynamics (Frick and Richter 1995). There is much discussion on whether the dynamical processes are defined by a spectrum of exponential relaxations (heterogeneous) or by a single intrinsically nonexponential process (Arbe *et al.* 1998).

The former possibility requires variation in the static structure to provide a variety of environments, which would be consistent with the presence of some local order. At present the evidence for both modes are equally weighted. Some polymer glasses show so-called boson peaks (related to the low energy vibrational density of states) that have been related to nanostructures in both polymethylmethacrylate (Duval *et al.* 1998) and polystyrene (Frick *et al.* 1995), which provides some support for heterogeneity both in the static structure and the relaxation behavior.

4. Polystyrene

Polystyrene is perhaps the mostly widely studied glassy polymer in terms of its short-range order.

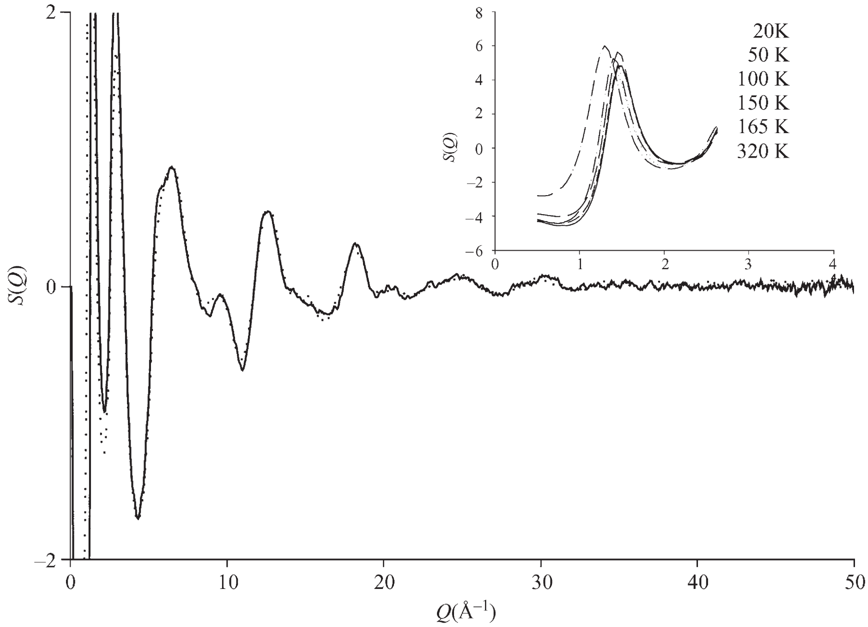


Figure 2
Neutron structure factors for deuterated *trans* 1,4 polybutadiene ($M_W \sim 10^5$, $M_W/M_n = 1.06$) measured at 20 K (solid line, glassy state) and at 320 K (broken line, melt state). The scale has been chosen to show the data at high *Q* values; the inset shows the low *Q* data for the temperatures shown. The glass transition for this material is ~160 K (after Gkourmpis and Mitchell 2000).

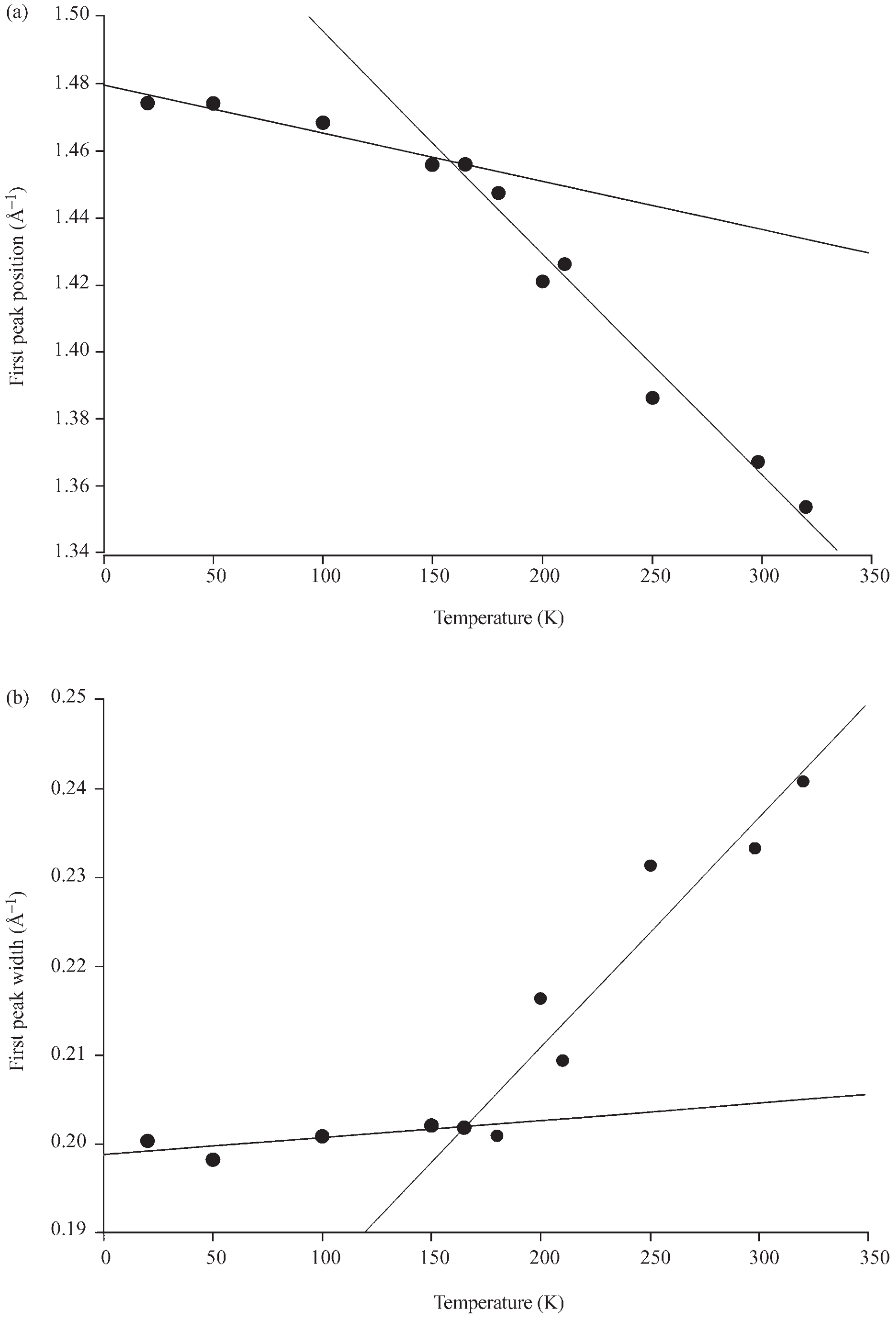


Figure 3
A plot of (a) the first peak position and (b) the first peak width as a function of temperature for the neutron scattering shown in Fig. 2. The solid lines represent linear fits to the data points either side of the glass transition temperature (after Gkourmpis and Mitchell 2000).

Commercial grades of polystyrene are usually formed from atactic polystyrene, although both isotactic and syndiotactic systems are now available and crystallization can be inhibited through modest quenching rates from the melt (for example Fig. 1). Studies of the local structure date from the work of Katz who identified a so-called polymerization peak in the x-ray scattering pattern of polystyrene that was not present in the equivalent pattern of liquid styrene (Katz 1936). Various studies have been made of this polymerization peak, which displays strong temperature dependence in terms of its position and intensity. Mitchell and Windle proposed a "superchain" model to account for the polymerization peak in which the pendant phenyl rings aggregate on a local scale to form stacks (Mitchell and Windle 1984).

NMR studies indicate that the mutual orientation of the phenyl rings is largely random (Robyr *et al.* 1995). The bulky side-groups dominate the broad Q scattering patterns and restrict the local conformational information that can be extracted from the data. One promising approach is to use broad Q neutron scattering coupled with selective isotopic substitution to yield partial structure factors corresponding to correlations between the skeletal atoms in the polymer chains (Mitchell 1995). Roe and co-workers found reasonable agreement between scattering functions derived from molecular simulations and experimental x-ray and neutron scattering data (Mondello *et al.* 1994, Furuya *et al.* 1994, Roe *et al.* 1995). The model was based on a cube of length ~24 Å and hence the longest length which can be considered with any confidence is half the cell size which is close to the distance corresponding to the polymerization peak. Detailed analysis of the "best" model showed that any parallelism between chain backbones was limited to separations <6 Å, while there was a limited degree of mutual alignment between neighboring phenyl rings.

Larger models (cube length ~39 Å) were generated by Kotelyanskii *et al.* (1996) using a combination of a lattice model and "decoration" using atomistic level segments but the authors conclude that even larger models are required to provide a realistic simulation of glassy polystyrene. One possible approach to generating larger models is the so-called reverse Monte Carlo procedures (McGreevy 1993, Rosi-Schwartz and Mitchell 1994) in which random adjustments are made to an atomistic level model and these moves are rejected or accepted not on the basis of some potential energy change, but through a cost function derived from the differences between the calculated and experimental scattering functions.

5. *Polymethylmethacrylate*

Atactic polymethylmethacrylate is widely used in optical components for a high level of clarity and structural rigidity. The intermediate level order of polymethylmethacrylate has been extensively studied since the steric crowding along the chain leads to a curved backbone conformation. These proposals have been confirmed using wide-angle x-ray scattering measurements (Lovell and Windle 1981 , Waring *et al.* 1982). These and more recent broad Q neutron scattering studies coupled with molecular modeling support the view of a persistent chain conformation in an otherwise highly disordered material (Ward and Mitchell 1995). Gabrys *et al.* (1986) reported the existence of some short-range order in glassy polymethylmethacrylate using spin polarization neutron scattering, although their contentions are not supported by molecular models (Gabrys *et al.* 1986). Miller *et al.* (1984) obtained wide-angle x-ray scattering from systematic series of amorphous acrylate and methacrylate polymers with differing alkyl side groups. The regular variation of the positions of the diffuse peaks in the scattering patterns with alkyl length led to the proposal of local ordered structures.

6. *Polycarbonate*

Polycarbonates are important thermoplastics combining optical clarity with good fracture toughness. In contrast to polystyrene and polymethylmethacrylate, polycarbonate is a crystallizable system, however long times (~30 days) are required for isothermal crystallization; the addition of a solvent can accelerate this process. There has been much speculation of the role of the local structure in determining the mechanical properties of polycarbonates particularly the embrittlement that takes place after substantial physical aging. The most commonly studied system is bisphenol A polycarbonate. Cervinka *et al.* proposed a model on the basis of broad Q neutron scattering in which the polycarbonate chains lie parallel (Cervinka *et al.* 1987). The same model was supported by data obtained by spin polarization neutron scattering techniques on both polycarbonate and some chemically modified carbonates (Lamers *et al.* 1994).

However, the authors emphasize that the models were small in size compared with the structural features present in polycarbonate. Eihard *et al.* (1999) used atomistic level mapping on to a coarse-grained model to generate a model with the correct large-scale chain configurations. They found good agreement between calculated and experiment neutron scattering structure factors. The possibility of some local order between phenyl groups and carbonate groups was identified by Robyr *et al.* using ^{13}C polarization transfer NMR but they emphasized the limited agreement between atomistic level simulation and experiment (Robyr *et al.* 1998) while other NMR measurements suggest that the level of local orientational correlations in the glass are similar to those present in the melt (Utz *et al.* 1999). Variations of structure with temperature in polycarbonates have been reported by del Val *et al.* (1995) and Mitchell and Windle (1985a), the latter propose that physical aging leads to an enhancement of local order which serves to inhibit crystallization through the presence of a multitude of equivalent nucleation sites.

Tucker *et al.* (2000) have used neutron scattering to study the variation in the local order of polycarbonate blended with an ionomer. They propose that the local order is destroyed on mixing. Mitchell and Windle (1985b) have demonstrated in a related polymer blend that mixing may lead to new structures again underpinning the notion of some local order.

7. Summary

By their nature broad Q x-ray and neutron scattering and NMR procedures emphasize the presence of well-defined distances in the local structural arrangements of glassy polymers. It would seem very reasonable that, on geometric grounds alone, different polymers would generate local structure which reflect the specific nature of the repeating units. Such arrangements, particularly at normal temperatures and pressures, may well represent the most disordered state possible. The presence of short-range order must require some correlations over and above those arising from simple geometric packing for even simple atomic fluids exhibit well-defined structure factors even though the packing units are spherical.

Overall there appears to be little evidence for such short-range order over and above that of the disordered state. However, polymers with well-defined local geometries such as polycarbonates and other aromatic polymers, together with systems exhibiting strong dipoles (for example PVC; Smith *et al.* 1993) show some characteristics of locally ordered structures. The increasing availability of high quality broad Q neutron scattering data and detailed information from NMR techniques when tightly coupled to atomistic level models may enable more precise information to be extracted. Realistic models of polymer glasses are vitally important for developing an understanding of many physical properties ranging from diffusion to fracture and yield.

Bibliography

Arbe A, Colmenero J, Monkenbusch M, Richter D 1998 Dynamics of glass-forming polymers: homogeneous versus heterogeneous scenario. *Phys. Rev. Lett.* **81**, 590–3

Cervinka L, Fischer E W, Hahn K, Jiang B Z, Hellman G P, Kuhn K L 1987 On the intermolecular order in amorphous polycarbonate—neutron scattering and model calculations. *Polymer* **28**, 1287–92

Clarke J H R, Brown D 1994 Molecular dynamics modeling of amorphous polymers. In: Colbourn E A (ed.) *Computer Simulation of Polymers*. Longman, Harlow, UK, pp. 46–90

del Val J J, Colmenero J, Rosi B, Mitchell G R 1995 X-ray diffraction study of the influence of temperature on the structural correlations of poly(2-hydroxypropyl ether of bisphenol A). *Polymer* **36**, 3625–31

Duval E, Mermet A, Surovtsev N V, Dianoux A J 1998 Boson peak, nanostructure, and relaxation of glasses. *J. Non-crystalline Solids* **235**, 203–7

Eihard J, Zirkel A, Tschoep W, Hahn O, Kremer K, Scharpf O, Richter D, Buchenau U 1999 Spatial correlations in polycarbonates: neutron scattering and simulation. *J. Chem. Phys.* **110**, 1819–30

Frick B, Buchenau U, Richter D 1995 Boson peak and fast relaxation process near the glass transition in polystyrene. *Colloid Polym. Sci.* **273**, 413–20

Frick B, Richter D 1995 The microscopic basis of the glass transition in polymers from neutron scattering studies. *Science* **267**, 1939–45

Furuya H, Mondello M, Yang H-J, Roe R-J, Erwin R W, Han C C, Smith S D 1994 Molecular dynamics simulation of atactic polystyrene. 2. Comparison with neutron scattering data. *Macromolecules* **27**, 5674–80

Gabrys B 1994 Short-range order in amorphous polymers. *TRIP* **1(2)**, 2–5

Gabrys B, Higgins J S, Schapf O 1986 Short-range order in amorphous polymethylmethacrylate. *J. Chem. Soc. Far. Trans. 1* **82**, 1929–34

Gkourmpis T, Mitchell G R 2000 Local order of polybutadiene melts and glasses. *J. Condensed Matter*

Higgins J S, Benoit H C 1994 *Polymers and Neutron Scattering*. Clarendon Press, Oxford

Katz J R 1936 X-ray diffraction of polymers. *Trans Faraday Soc.* **32**, 77–89

Kotelyanskii M, Wagner N J, Paulaitis M E 1996 Building large amorphous polymer structures: atomistic simulation of glassy polystyrene. *Macromolecules* **29**, 8497–506

Lamers C, Schonfield C, Shapiro S M, Batoulis J, Timmermann R, Cable J W, Richter D 1994 Short-range order effects in amorphous polycondensates as studied by spin polarization diffuse neutron scattering and simulation. *Colloid Polym. Sci.* **272**, 1403–9

Lovell R, Windle A H 1981 Determination of the local conformation of PMMA from wide-angle x-ray scattering. *Polymer* **22**, 175–84

McGreevy R L 1993 Neutron diffraction and RMC modeling—complementary techniques for the determination of disordered structures. *Int. J. Mod. Phys. B* **7**, 2965–80

Miller R L, Boyer R F, Heijboer J 1984 X-ray scattering from amorphous acrylates and methacrylates: evidence for local order. *J. Polym. Sci. Polym. Phys. Edn.* **22**, 2021–41

Mitchell G R 1985 The local structure on noncrystalline polymers: an x-ray approach. In: Keinath S E, Miller R L, Reike J R (eds.) *Order in the Amorphous "State" of Polymers*. Plenum, New York, pp. 1–31

Mitchell G R 1989 X-ray scattering from noncrystalline and liquid crystalline polymers. In: Allen G, Bevington J C, Booth C, Price C (eds.) *Comprehensive Polymer Science*. Pergamon, Oxford, Vol. I, Chap. 31, pp. 687–729

Mitchell G R 1995 Local order in polymer glasses and melts. In: Keller A, Warner M, Windle A H (eds.) *Self-order and Form in Polymeric Materials*. Chapman and Hall, London, Chap. 7, pp. 95–113

Mitchell G R, Windle A H 1984 Structure of polystyrene glasses. *Polymer* **25**, 906–20

Mitchell G R, Windle A H 1985a The effect of annealing on the glassy structure of polycarbonate. *Colloid Polym. Sci.* **263**, 280–5

Mitchell G R, Windle A H 1985b The local structure of blends of polystyrene and poly(2,6-dimethylphenylene). *J. Polym. Sci. Polym. Phys. Edn.* **23**, 1967–74

Mondello M, Yang H-J, Furuya H, Roe R-J 1994 Molecular dynamics simulation of atactic polystyrene. 1. Comparison with x-ray scattering data. *Macromolecules* **27**, 3566–74

Monnerie L, Suter U W 1994 *Advances in Polymer Science 116. Atomistic Modeling of Physical Properties*. Springer-Verlag, Berlin

Robyr P, Gan Z, Suter U W 1998 Local order between chain segments in the glassy polycarbonate of 2,2-bis(4-hydroxyphenyl)propane from ^{13}C polarization-transfer NMR. *Macromolecules* **31**, 6199–205

Robyr P, Tomaselli M, Grob-Pisano C, Meier B H, Ernst R R, Suter U W 1995 Characterization of local order in atactic polystyrene using two-dimensional nuclear magnetic resonance and atomistic simulations. *Macromolecules* **28**, 5320–4

Roe R-J 2000 *Methods of X-ray and Neutron Scattering in Polymer Science*. Oxford University Press, New York

Roe R-J, Mondello M, Furuya H, Yang H J 1995 Molecular dynamics simulation of atactic polystyrene. 3. Short-range order. *Macromolecules* **28**, 2807–18

Rosi-Schwartz B, Mitchell G R 1994 A complete atomistic model of molten polyethylene from neutron scattering data—a new methodology for polymer structure. *Polymer* **35**, 5398–407

Schmidt-Rohr K, Spiess H W 1994 *Multidimensional Solid-state NMR and Polymers*. Academic Press, London

Smith G D, Jaffe R L, Yoon D-Y 1993 Conformations and order in atactic poly(vinyl chloride) melts from molecular dynamics simulations. *Macromolecules* **26**, 298–304

Tucker R, Gabrys B, Zajac W, Andersen K, Kalhoro M S, Weiss R A 2000 Destruction of short-range order in polycarbonate-ionomer blends. In: Cebe P, Hsiao B S, Lohse D J (eds.) *ACS Symposium Series 739*. American Chemical Society, Washington, DC, Chap. 21 , pp. 328–40

Utz M, Atallah A S, Robyr P, Widmann A H, Ernst R R, Suter U W 1999 Solid-state NMR investigation of the structural consequences of plastic deformation in polycarbonate. 1. Global orientational order. *Macromolecules* **32**, 6191–205

Ward D J, Mitchell G R 1995 Modeling the structure of a polymer glass polymethylmethacrylate through neutron scattering experiments. *Physica Scripta* **T57**, 153–60

Waring J R, Lovell R, Mitchell G R, Windle A H 1982 Radial distribution functions of noncrystalline polymers and their application to the structural analysis of polymethylmethacrylate. *J. Mater. Sci.* **17**, 1171–86

Wendorff J H 1985 Interrelationships between molecular and physical structure in amorphous polymers. In: Keinath S E, Miller R L, Reike J R (eds.) *Order in the Amorphous "State" of Polymers*. Plenum, New York, pp. 53–70

G. R. Mitchell
University of Reading, UK

T

Teeth: Structure/Mechanical Design Strategies

Teeth fulfill a function that is almost always essential for survival—food procurement and processing. This constant evolutionary pressure ensures that their structures are finely tuned to their function. Furthermore, any benefits that might result from natural selection will readily be incorporated into the structure. Thus from the materials perspective, teeth can be regarded as cutting, piercing, and grinding tools whose structures and shapes have been finely honed over tens and even hundreds of millions of years. An understanding of the structure–function relations of teeth may therefore provide new ideas for the design of improved cutting, piercing, and grinding tools, and have implications for the development of novel dental and orthopedic materials.

A wide variety of animals produce teeth, not all of which are mineralized. Presumably, the working surfaces of all, however, need to be hardened in one way or another to function well. The remainder of the tooth appears to provide a backing for the working surface. Thus, the very basic design strategy of most teeth seems to follow this intuitive division. Judging from the complexity of tooth structure, however, there are many more functional design features in teeth. Here we will briefly consider several tooth types whose structure–function relations have been investigated. For more information on mechanical properties of biological materials see Wainwright *et al.* (1976), and on biomineralization see Lowenstam and Weiner (1989) and Simkiss and Wilbur (1989).

1. Sea Urchin Teeth

Sea urchins (phylum Echinodermata) produce five continuously growing teeth (Fig. 1(a)) that are used for grinding the rocky substrate in order to extract the algae that tightly adhere to the surface. In fact, some of the rock itself is ground away. Amazingly, the tooth is composed of basically the same material as the rock that they usually grind, namely calcium carbonate. The tooth structure and modifications of the mineral phase are what makes it possible to grind away the rocky substrate. The tooth is continuously renewed as the scraping surface is worn down, and is self-sharpening (Fig. 1(a)) (Hyman 1955). For more information on the overall tooth structure see Markel *et al.* (1977) and Stock *et al.* (2002). Here we will consider only a few of the structural features.

The basic structural elements of the tooth are long S-shaped fibers each of which is a single crystal of calcite. Each fiber is enveloped in an organic sheath. The fibers taper from 20 μm in diameter in the keel area to less than 1 μm or so in the working surface (Fig. 2(c)). There is also a reduction in the magnesium content from the keel side to the working surface (Wang *et al.* 1997). The reduction in size results in the hardness of the element increasing due to the decrease in the number of possible dislocations. The working surface (Fig. 2(b)) is composed of the tips of the fibers emerging perpendicular to the surface. Each fiber is embedded in a matrix of rounded and somewhat disordered crystals of disk-shaped calcite that contain very large amounts (up to 35 mol.%) of magnesium within its lattice. The two phases are separated by an organic sheath (Wang *et al.* 1997). Resistance to fracture is presumably derived from the

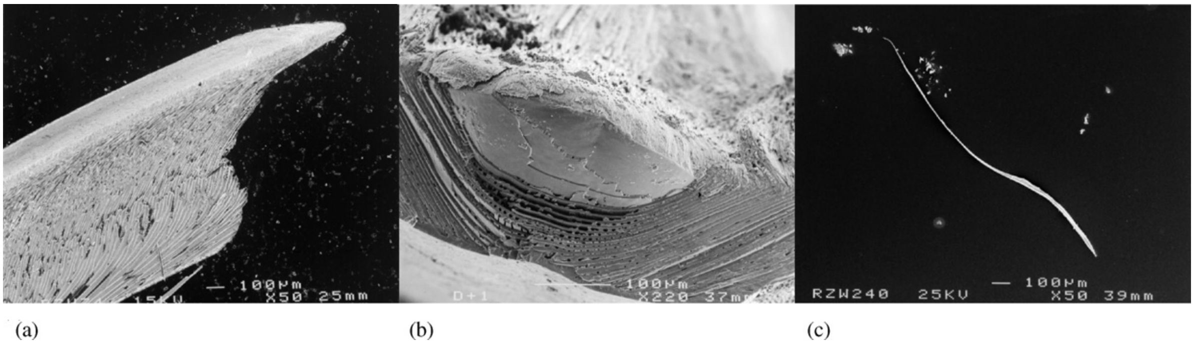


Figure 1
SEMs of a sea urchin (*Paracentrotus lividus*) tooth. (a) The working tip of the whole tooth showing the self-sharpening profile. (b) The working tip showing the core composed of calcite needles embedded in a matrix of high-Mg calcite crystals. The core is surrounded by plates. (c) A crystal fiber that tapers in size from one end to the other. The thin end is embedded in the working tip. Micrographs are adapted from those in Wang *et al.* (1997).

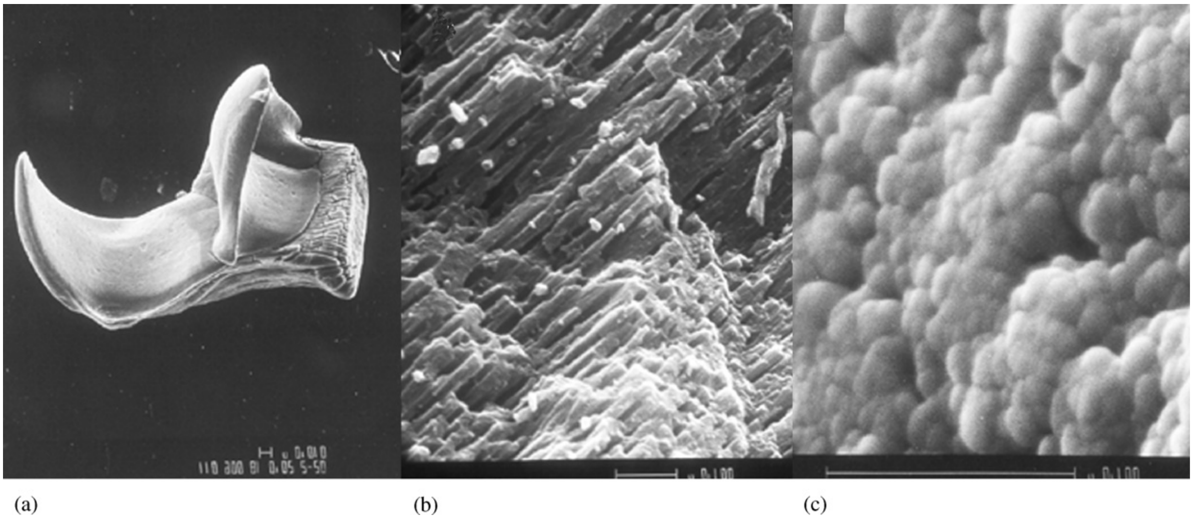


Figure 2
Micrographs of the teeth of the limpet *Lottia* sp. (a) SEM of a mineralized tooth. (b) SEM of the hard outer working layer containing goethite in an organic matrix. (c) SEM of the softer inner layer containing opal in an organic matrix. Micrographs are from the collection of the late Prof. Heinz A. Lowenstam.

juxtaposition of the hard needles in a matrix of small rounded crystals. Furthermore, movement between the two can occur along the interface. In certain respects the tooth can be regarded as a gradient fiber-reinforced composite material in which both the fibers and the matrix are composed of mineral. For more information on structure and other design strategies built into the tooth structure, including the self-sharpening property, see Wang *et al.* (1997).

2. *Chiton and Limpet Teeth*

Chitons (phylum Mollusca, class Polyplacophora) and limpets (phylum Mollusca, class Gastropoda) also produce teeth for grinding the rocky substrate in order to extract the algae. Chitons and limpets wear down their teeth at the rate of approximately a row a day, and thus also produce new tooth rows at the same rate (Lowenstam 1962b). This occurs on a long tongue-like structure called a radula, which comprises more than 100 rows of teeth. Figure 2(a) shows a scanning electron micrograph (SEM) of a mineralized tooth of a limpet. The outer working layer is composed of goethite (iron hydroxide mineral) in an organic matrix (Fig. 2(b)), and the remainder of the tooth is composed of amorphous silica (opal) in an organic matrix (Fig. 2(c)) (Lowenstam 1962a, 1971).

In chitons only two of the teeth in each row are mineralized. The mineral of the working surface layer is the relatively hard iron oxide magnetite—in fact the teeth are magnetic. Lowenstam (1962b) was the first to identify biogenic magnetite in the chiton tooth, or for that matter in any organism. It is now known to be widespread in biology (Kirschvink *et al.* 1985). The remainder of the tooth is comprised of softer and more pliable mineral phases, which in certain species is carbonated apatite (the same mineral present in vertebrate bones and teeth) or in other species an amorphous calcium iron phosphate mineral (Lowenstam and Weiner 1989). The two layers are separated by a thin third layer also containing an amorphous mineral, ferrihydrite. This third layer possibly functions as a gasket allowing the two major layers with distinctly different mechanical properties to "work" together—a supposition that needs to be checked in chiton teeth, but has supporting evidence in human teeth (see below).

van der Wal (1989) characterized the structure and materials design properties of the mineralized teeth of a limpet. Of particular interest is the manner in which the crystals in the working tip have two different orientations, and the hardness of the tip is about twice that of the other parts of the tooth. These properties are the key to the functioning of the tip as a self-sharpening grinding tool.

3. *"Jaws" of Carnivorous Worms*

Various marine polychaete worms have sets of jaws which are a few millimeters long and sharp (Fig. 3). The tip is used by the worm as a syringe to inject venom into its prey. Bryan and Gibbs (1979) reported high levels of zinc in some species and copper in others. Lichtenegger *et al.* (2002) showed that in the case of the copper-containing species, *Glycera*, some of the copper is in the form of a most unusual mineral phase called atacamite ($Cu_2(OH)_3Cl$). The atacamite crystals are located in the region of the tips of the jaws. There are four polymorphs of this mineral phase, and the animals somehow induce the formation of only one (Lichtenegger *et al.* 2002).

Lichtenegger *et al.* (2002) performed nanoindentation measurements as well as microprobe elemental analyses. They mapped the hardness and the elastic modulus, and found that the areas with higher hardness and moduli are indeed those with high copper concentrations. In fact, the values measured were intermediate between dentin and enamel, the two materials of vertebrate teeth (see below). They also estimated the resistance of the worm teeth material to abrasion, and found that it is twice that of dentin and almost 80% of that of enamel, despite the fact that amount of mineral in the tissue is just one-seventeenth of that in dentin.

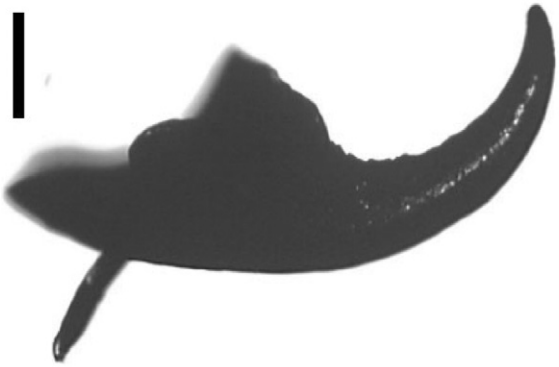


Figure 3
Photograph of the jaw showing the sharp point used for injecting the venom. Scale: 0.2 mm (courtesy of Helga Lichteneggar).

Lichtenegger *et al.* (2002) also noted that the distribution of copper is higher in the inner part of the jaw, as compared to the surface. A gradual increase of stiffness from the surface to the interior is known to prevent crack formation and improve the ability of stiff materials to resist contact damage (Suresh *et al.* 1999). Thus, the jaw design strategy integrates protein, copper ions, and copper minerals, and the authors point out that this may serve as a design prototype for new materials.

4. *Human Teeth*

Human teeth, like most teeth of vertebrates, are composed of a hard thin working surface (enamel) that overlies the bulk of the tooth which is composed of a softer more pliant material called dentin (Fig. 4(a)). The mineral phase in both layers is composed of carbonated apatite (also known as dahllite), but the crystal sizes and shapes are very different. In enamel the crystals are highly elongated (spaghetti-shaped). They are arranged in bundles or prisms, and in many species, these bundles are aligned in no less than three different orientations (Fig. 4(b)). The mineral phase comprises ~99% by weight of the material. The dentin, on the other hand, contains exceedingly small crystals (among the smallest known to be formed biologically). They are plate-shaped with a thickness of just 2–4 nm. The mineral comprises ~65% by weight. A major portion of the dentin crystals are located inside fibrils of collagen, the protein that comprises the bulk of the matrix. They are arranged in layers that traverse across the fibril (reviewed in Lowenstam and Weiner 1989).

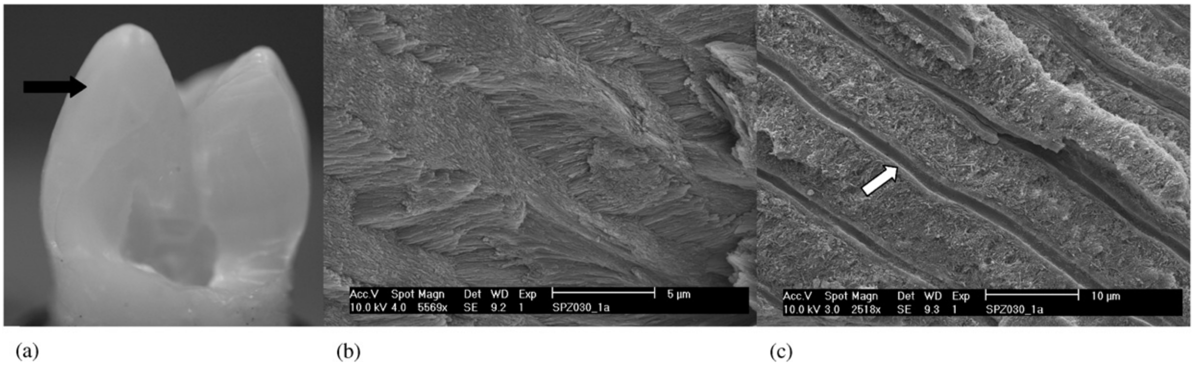

Acc.V Spot Magn Det WD Exp 5 µm
10.0 kV 4.0 5569x SE 9.2 1 SPZ030_1a
Acc.V Spot Magn Det WD Exp 10 µm
10.0 kV 3.0 2518x SE 9.3 1 SPZ030_1a
(a)
(b)
(c)

Figure 4
Micrographs of human teeth. (a) View of a fractured cross-section of a human premolar. The arrow shows the outer enamel layer. (b) SEM of human enamel showing parts of prisms with elongated crystals oriented in different directions. (c) SEM of dentin in the crown of the tooth, showing the presence of peritubular dentin around the tubules (arrow).

The dentin and, to a certain extent, the enamel are graded materials in that their structural and hence mechanical properties continuously change. This has been demonstrated by microhardness profiles (Craig *et al.* 1959). One characteristic of dentin is the presence of numerous long and thin tubules (Fig. 4(c)). They usually have an internal diameter of 1 μm and extend from just below the dentin–enamel junction (DEJ) to the pulp cavity. The dentin immediately below the enamel layer in the crown of the tooth contains yet another material type called peritubular dentin. This is a dense collar of mineral that surrounds each of the tubules (Fig. 4(c)). It contains crystals of the same shape, size, and basic organization as in the bulk of the dentin, but almost no collagen (Weiner *et al.* 1999). It is much harder than intertubular dentin and presumably much stiffer.

Various materials properties of enamel and dentin have been measured separately. These include both elastic and fracture properties (Kinney *et al.* 1996). Surprisingly, very little is known about the design strategy of whole teeth. Some human teeth (canines for example) function mainly as cutting tools, whereas others are mainly for grinding. During mastication the forces applied to the tooth are not only exerted in a direction perpendicular to the cutting surface, but can be from oblique directions. Furthermore, the manner in which the stress is applied can vary greatly depending upon the nature of the food being chewed. Thus, the tooth structure presumably has many design strategies for dealing with different mechanical challenges. Here we identify a few.

The position and orientations of the tubules surrounded by peritubular dentin location below the enamel working surface point to a buttressing function for peritubular dentin such that it supports the enamel cap. The tubules with their collar of hard peritubular dentin provide added stiffness to the crown dentin. Strangely, the peritubular dentin does not extend all the way to the dentin–enamel interface or junction (DEJ). Nor do the tubules extend all the way to the DEJ. Both are absent in a transition zone some 100–200 μm below the DEJ. In this same zone, the hardness decreases significantly as compared to the bulk dentin below. All these structural properties point to a unique role for this area of the dentin (Weiner *et al.* 1999).

Wang and Weiner (1998) reasoned that because of the marked differences in stiffness between the enamel and the dentin, some sort of soft intermediate layer must be present to allow the two materials to "work" together. The microhardness profiles supported this supposition. They used Moiré interferometry to map the strain on the surface of a human tooth slice subjected to compressive forces comparable to those incurred during mastication. They found that indeed the highest strains were measured in the 200 μm zone below the DEJ. Furthermore, the extent of the strain was asymmetric when comparing the labial (external) and the lingual (internal) sides of the tooth. It was postulated that this soft zone is actually the working part of the tooth (Wang and Weiner 1998). It apparently acts as a gasket to allow the very hard and stiff enamel to "work" together with the softer more pliant dentin. Similar experiments were repeated at much higher resolution, but instead of using compressive forces, the tooth slices were subjected to changes of humidity. In one experimental configuration the enamel confined the dentin and prevented it from contracting as a function of dehydration, whereas in another configuration the dentin contracted and expanded freely. Most of the expansion was in the same soft zone below the DEJ (Wood *et al.* 2000).

Much more remains to be understood about the manner in which whole human teeth are designed to fulfill their functions. This information may have far-reaching implications on treatment regimes and on the design of new dental materials.

5. Concluding Comment

Teeth structures throughout the animal kingdom appear to have some common design features, but in general are very different. All, however, probably fulfill vital functions and, therefore, have interesting structure–function relations. Much can be learned from the study of these relations that may produce new ideas for the design of improved synthetic materials.

Bibliography

Bryan G W, Gibbs P E 1979 Zinc—a major inorganic component of Nereid polychaete jaws. *J. Mar. Biol. Ass. UK* **59**, 969–73

Craig R, Gehring P, Peyton F 1959 Relation of structure to the microhardness of human dentin. *J. Dent. Res.* **38**, 624–30

Hyman L 1955 *The Invertebrates. IV. Echinodermata*. McGraw-Hill, New York

Kinney J, Balooch M, Marshall S, Marshall G, Weihs T 1996 Hardness and Young's modulus of human peritubular and intertubular dentine. *Arch. Oral Biol.* **41**, 9–13

Kirschvink J L, Jones D S, McFadden B J 1985 *Magnetite Biomineralization and Magnetoreception in Organisms*. Plenum, New York and London

Lichtenegger H, Schoberl T, Bartl M, Waite H, Stucky G 2002 High abrasion resistance with sparse mineralization: copper biomineral in worm jaw. *Science* **298**, 389–92

Lowenstam H A 1962a Goethite in radular teeth of recent marine gastropods. *Science* **137**, 279–80

Lowenstam H A 1962b Magnetite in denticle capping in recent chitons (Polyplacophera). *Geol. Soc. Am. Bull.* **73**, 435–8

Lowenstam H A 1971 Opal precipitation of marine gastropods. *Science* **171**, 487–90

Lowenstam H A, Weiner S 1989 *On Biomineralization*. Oxford University Press, New York

Markel K, Gorny P, Abraham K 1977 Microarchitecture of sea urchin teeth. *Fortschr. Zool.* **24**, 103–14

Simkiss K, Wilbur K 1989 *Biomineralization. Cell Biology and Mineral Deposition*. Academic Press, San Diego

Stock S, Dahl T, Barss J, Veis A, Fezzaa K, Lee W 2002 Mineral phase microstructure in teeth of the short spined sea urchin (Lytechinus variegatus) studied with X-ray phase contrast imaging and with absorption microtomography. *Adv. X-ray Anal.* **45**, 133–8

Suresh S, Olsson M, Giannakopoulos A E, Padture N P, Jitcharoen J 1999 Engineering the resistance to sliding-contact damage through controlled gradients in elastic properties at contact surfaces. *Acta Mater.* **14**, 3915–26

van der Wal P 1989 Structural and material design of mature mineralized radula teeth of Patella vulgata (Gastropoda). *J. Ultrastruct. Mol. Struct. Res.* **102**, 147–61

Wainwright S A, Biggs W D, Currey J D, Gosline J M 1976 *Mechanical Design in Organisms*. Princeton University Press, Princeton

Wang R, Weiner S 1998 Strain–structure relations in human teeth using Moiré fringes. *J. Biomech.* **31** (2), 135–41

Wang R Z, Addadi L, Weiner S 1997 Design strategies of sea-urchin teeth—structure, composition and micromechanical relations to function. *Phil. Trans. R. Soc. Ser. B* **352** (1352), 469–80

Weiner S, Veis A, Beniash E, Arad T, Dillon J, Sabsay B, Siddiqui F 1999 Peritubular dentin formation: crystal organization and the macromolecular constituents in human teeth. *J. Struct. Biol.* **126**, 27–41

Wood J D, Wang R Z, Weiner S, Pashley D H 2000 Mapping of tooth deformation caused by moisture change using Moiré interferometry. *Dent. Mater.* **19**, 159–66

S. Weiner and P. Zaslansky
Weizmann Institute of Science, Rehovoth, Israel

Tissue Engineering Scaffolds

Tissue engineering scaffolds fulfill three primary purposes. First, they form regenerate tissue shape. Second, they temporarily fulfill the physical function (mechanical, chemical, and/or electrical) of the native tissue. And finally, their material composition and porous microstructure enhance tissue regeneration. Many studies have noted that scaffold materials and pore morphology significantly influence regeneration of either bone or cartilage (Sampath and Reddi 1984, Kuboki *et al.* 1995, 1998, Tsuruga *et al.* 1997). Furthermore, pores with appropriate diameter alone are not sufficient for bone regeneration; the pores must also have adequate connectivity and orientation for the regenerated bone tissue to carry mechanical loads. Bruder *et al.* (1998) noted that despite a large amount of regenerated bone tissue, calcium phosphate scaffolds with uncontrolled pore architecture produced poor biomechanical properties. Like Bostrom and Mikos (1996), they have demonstrated that controlling internal scaffold architecture is crucial to controlling the vascularization and mechanical properties of regenerate tissue. Moreover, scaffolds must also replicate complex external shapes. Therefore, the ability to control scaffold architecture, material composition and porosity through design and fabrication could be a critical factor in the future clinical success of tissue engineering. An ideal for tissue engineering should thus be to combine design and automated manufacturing techniques to enable patient specific bioimplants and tissue reconstruction strategies.

Polymers based on polylactic acid (PLA), polyglycolic acid (PGA) and their co-polymers (PLGA) are the most widely investigated biocompatible, biosorbable polymers for such bioimplants (Rothen-Weinhold *et al.* 1999, Stancari *et al.* 2000) and controlled release drug delivery devices (Putney 1999). Composites of these polymers with hydroxyapatite (HA) and calcium phosphate are being investigated as bone fixation devices in orthopedic, craniofacial, maxillofacial, and reconstructive surgeries. Polymethyl-methacrylate (PMMA), Nylon-6 (Risbud and Bhonde 2001), Polycaprolactone (Zein *et al.* 2002), and bioactive glass (Stamboulis *et al.* 2002) are also being investigated.

To date, fabrication methods have demonstrated scaffolds in monolithic polymers or polymer-ceramic composites with limited control over geometry and porosity. Although significant research has been performed on the biologic (Caplan and Bruder 1997, Reddi 1998, Krebsbach *et al.* 1997) and material fabrication (Bostrom and Mikos 1996, Yazemski *et al.* 1996) aspects of bone tissue engineering scaffolds, controlled three-dimensional scaffold design and fabrication have only recently received attention. The automated design of scaffolds with simultaneous control over external shape and internal architecture requires the use of computational design techniques akin to those developed by Hollister *et al.* (1999). These techniques enable rapid design and analysis of three-dimensional scaffold topologies that can also mimic natural bone structure. However, scaffold topologies generated by these computational techniques for the surgical reconstruction of complex defects present significant manufacturing challenges due to their complexity. Solid freeform fabrication (SFF) techniques have been successful at manufacturing complex scaffolds with better architecture control, e.g., to make bone tissue replacements from monolithic degradable biomaterials. In the future, SFF will be one of the most promising ways to manufacture heterogeneous scaffolds with graded compositions that can potentially regenerate multiple types of tissues or tissue interfaces simultaneously.

1. Computational Design of Scaffolds

Scaffold design requirements must be addressed on both macroscopic and microscopic scales. On a macroscopic scale (defined as >1 mm), a scaffold design must replicate human anatomy while on a microscopic level, it must fulfill temporary tissue function and simultaneously enhance tissue regeneration. These two scales must be integrated to produce a single design that can be embodied in a format appropriate for SFF methods.

Data for macroscopic design is typically derived from computed tomography (CT) or magnetic resonance imaging (MRI) that represents anatomical shape and topology by density distributions within a three-dimensional volume element (voxel) database. Image processing techniques, including basic thresholding, noise filtering, edge detection, morphological filtering, and region of interest selection, are needed to process the data into a form suitable for design purposes. Once the image data is processed, there are two ways to utilize it for scaffold design. For traditional computer aided design (CAD) techniques the image data must first be converted to a geometric form such as points, lines or surfaces. Geometric representations of anatomy can then be further manipulated to control the exterior scaffold shape. A second method directly creates topology within the voxel data structure, a process known as image-based design (IBD). This technique creates an external design by altering the density distribution within the voxel dataset.

Requirements for microscopic design are more complex and much less defined than macroscopic requirements. Both physical properties (including stiffness, permeability and conductivity) and tissue regeneration enhancement (related to scaffold porosity, pore diameter, interconnectivity, and permeability) depend directly on scaffold microstructure. Although these general requirements are agreed upon (LeBaron and Athanasiou 2000, Butler *et al.* 2000, Hollister *et al.* in press, Yang *et al.* 2001), the most critical requirement for successful tissue regeneration is yet to be determined, but must entail the concurrent design of scaffold microstructure, material, and tissue interface. A widely proposed design principle, although lacking substantive experimental evidence, is that scaffold microstructure should render effective properties that match those of native tissue and yet maintain high permeability or porosity for biofactor delivery. Three design approaches have been employed in accordance with this principle:

(i) material process-defined design,
(ii) periodic cell-based design, and
(iii) biomimetic design.

In material process-defined design, the scaffold microstructure is completely determined by specific material processing techniques. Salt leaching techniques, commonly used to fabricate porous polymers, form microstructures that cannot be set *a priori* given little or no control over channel direction and interconnectivity. In periodic cell-based design, a unit cell with specific microstructure is repeated to create an entire scaffold. This technique can be used in combination with topology optimization methods to design microstructures that produce desired effective physical properties. Finally, in biomimetic design, scaffolds mimic natural tissue structure based on the notion that they will provide the best function and tissue regeneration. Ideally, biomimetic designs seek to replicate all aspects of tissue structure and function, but are the most difficult to achieve due to complexity of biological tissues, which are composites organized in a hierarchy from nanometers to millimeters. In addition, even if a scaffold matches native tissue structure, regenerate tissue that grows into its pore architecture may not match native tissue function.

1.1 Periodic Cell-based Design

Our research efforts have encompassed all three design approaches, concentrating on periodic cell-based and biomimetic design techniques to create scaffolds that meet the conflicting requirements of matching desired functional properties while maintaining high porosity. In the periodic cell-based design approach, homogenization theory (Hollister and Kikuchi 1994, Hollister *et al.* in press) can be used to compute a functional dependence of effective scaffold property $C^{\text{eff scaf}}$ and effective regenerate tissue property $C^{\text{eff regen tissue}}$ on scaffold base property C^{scaf}, scaffold microstructure M^{scaf}, regenerate tissue base property $C^{\text{regen tissue}}$ and regenerate tissue microstructure $M^{\text{regen tissue}}$. In this approach, the scaffold unit-cell microstructure comprises intersecting porous cylinders characterized by three pore diameters d_1, d_2, and d_3, shown in Fig. 1.

The dependence of effective scaffold property on scaffold base property and scaffold microstructure can then be written as:

$$\begin{aligned} C^{\text{eff scaf}} &= C^{\text{scaf}} M^{\text{scaf}}(d_1, d_2, d_3) \\ C^{\text{eff regen tissue}} &= C^{\text{regen tissue}} M^{\text{regen tissue}}(d_1, d_2, d_3) \end{aligned} \tag{1}$$

Equation (1) can be used to optimally design the effective scaffold stiffness and the effective regenerate tissue stiffness. The following optimization formulation designs scaffold pore geometry such that the effective scaffold stiffness and the effective regenerate tissue stiffness match native tissue stiffness as closely as possible while meeting constraints on the base scaffold material stiffness, on scaffold porosity, and on pore diameters.

Objective function:

$$\underset{C^{\text{scaf}}, d_1, d_2, d_3}{\text{Min}} \left\{ \begin{array}{c} \displaystyle\sum_{i=1}^{n=1\to 9} \left(\frac{C_i^{\text{eff regen tissue}} - C_i^{\text{eff native tissue}}}{C_i^{\text{eff native tissue}}} \right)^2 + \\ \displaystyle\sum_{i=1}^{n=1\to 9} \left(\frac{C_i^{\text{eff scaf}} - C_i^{\text{eff native tissue}}}{C_i^{\text{eff native tissue}}} \right)^2 \end{array} \right\}$$

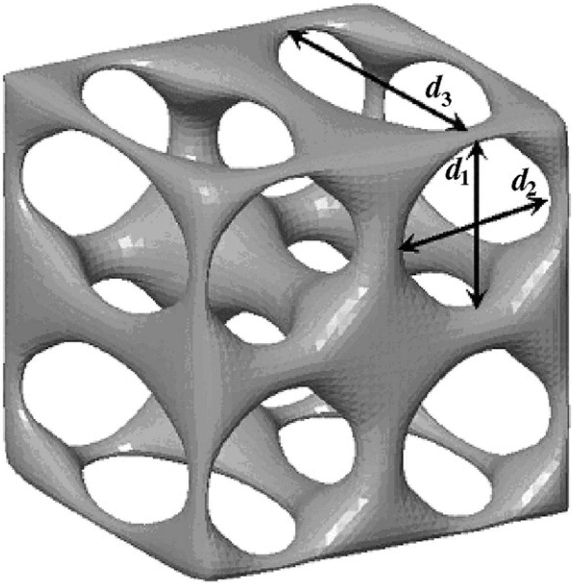


Figure 1
Example of periodic interconnecting porous microstructure used as the basic unit cell for optimizing scaffold microstructure.

Constraints:

$$
\begin{aligned}
&d_1,\ d_2,\ d_3 \le \text{largest pore size for regeneration}\\
&d_1,\ d_2,\ d_3 \ge \text{resolution of SFF technique}\\
&\frac{V_{\text{pore}}}{V_{\text{total}}} \ge \%\ \text{Porosity}\\
&C^{\text{scaf}} \ge C_{\min}\\
&C^{\text{scaf}} \le C_{\max}
\end{aligned}
\qquad (2)
$$

The optimization scheme in Eqn. (2) has produced interconnected cylindrical designs such that both the scaffold and the regenerate tissue match native bone stiffness with $R^2 = 0.99$. This optimization approach can also be used to design microstructures with functionally graded stiffness by varying pore size in all three dimensions, and by varying base scaffold material simultaneously.

1.2 Biomimetic Design

Biomimetic design requires the use of microstructural image data as a template. This imaging data may be microcomputed tomography data (micro-CT), micro-MRI or confocal microscopy data. We have used human trabecular bone architecture as the basis for creating biomimetic scaffold microstructures. Figure 2 shows a rendering of trabecular bone micro-CT data and the associated triangular facet data suitable for SFF. Such faceted data have been used to create scaffolds from Nylon-6 and Polylactic acid. Biomimetic designs present additional obstacles since the scale of biological tissue artifacts is often below the feature resolution of most freeform fabrication systems. However, these features can be scaled up for fabrication while retaining porosity sizes optimal for tissueregeneration.

2. SFF Methods for Tissue Engineering Scaffolds

SFF is a group of emerging manufacturing technologies whose common feature is an ability to produce freeform, geometrically complex objects directly from their computer-generated models (Beaman *et al.* 1997). A computer-generated (CAD) model of the object is mathematically sliced and each slice is then used to guide layer-by-layer material deposition in selected regions to build a complete, three-dimensional object. Recently, a number of researchers have studied the construction of tissue engineering scaffolds using SFF methods such as three-dimensional printing (Giordano *et al.* 1996, Cima *et al.* 1996a, 1996b, Vacanti *et al.* 2001), mechanical assembly of pre-fabricated layers (Weiss and Calvert 2000), fused deposition modeling (Zein *et al.* 2002), thermojet printing, ceramic stereolithography (Chu *et al.* 1999), and selective laser sintering. In the future, major advances in SFF are expected to take place in the development of machines, materials and techniques that enable the fabrication of heterogeneous, multifunctional biomedical devices.

2.1 Fabrication of Heterogeneous Scaffolds

The temporo-mandibular joint (TMJ) is among the most complex joints in the human body and ~100 000 Americans undergo surgical reconstruction of this joint annually. In order to reconstruct a whole TMJ, novel methods are needed to build complex, three-dimensional scaffolds incorporating multiple biomaterial and porosity gradients that upon implantation will enable the simultaneous growth and regeneration of multiple tissues (bone, cartilage, ligament, or tendon) and blood vessels. SFF methods are particularly well-suited to building such geometrically complex, heterogeneous structures because of the manner in which they construct objects — depositing multiple materials where needed; volume element by volume element; layer by layer. This fabrication capability when combined with computational design techniques for scaffolds with spatially distributed materials and properties is a very powerful combination. Using these methods, a whole joint structure including a bone scaffold (made from a ceramic or ceramic-polymer composite), combined with a cartilage scaffold (made with PGA) and a ligament

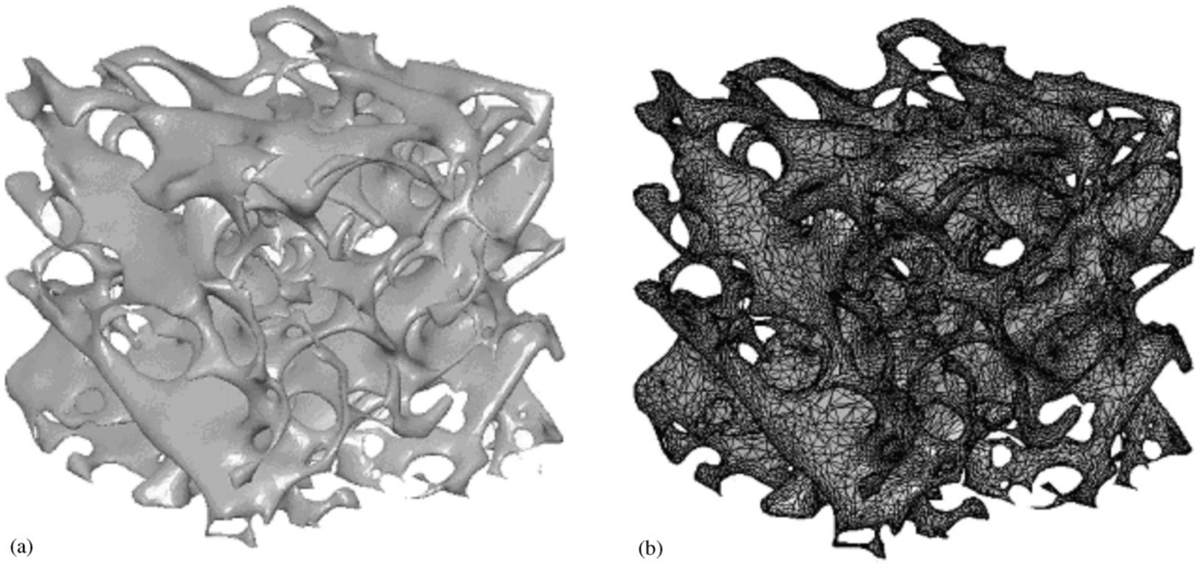


Figure 2
Trabecular bone microstructure used for biomimetic design. (a) Shaded rendering of image data. (b) Triangular facet data for an stereolithography file.

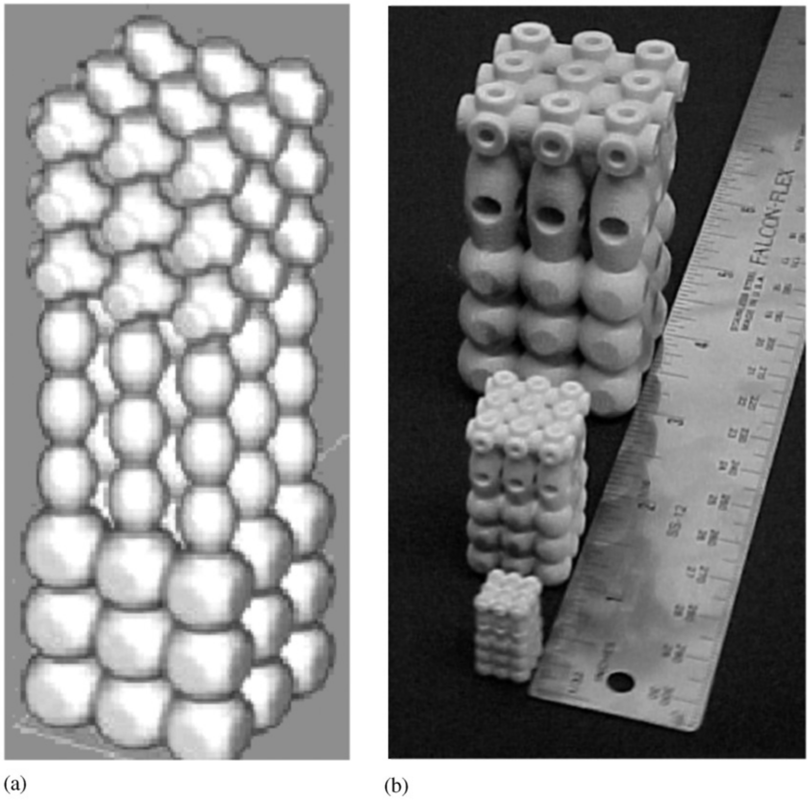


Figure 3
(a) Image based design of scaffold with functionally graded architecture. (b) Scale models constructed using SLS.

scaffold (made with PLGA) with different functional architectures could be built in a single assemblage. During construction, precise amounts of bioactive factors could also be incorporated into the scaffold.

An example of a scaffold with functionally graded architecture and porosity derived from image-based design methods is shown in Fig. 3(a). This structure transitions from a morphology of three-dimensionally connected spheres to ellipsoids connected only in the vertical direction, finally to orthogonally connected tubular channels. Figure 3(b) shows scale models we have constructed in Nylon–11 using selective laser sintering (SLS). When physically constructed using multiple, graded biocompatible, biosorbable polymers, and polymer-ceramic composites, this geometry (300–1000 μm typical feature size) will produce a gradient in material properties and porosity that could be used as a scaffold for engineered tissue interfaces.

In SLS, a laser-based SFF technique being investigated by the authors, an object is built by fusing sequentially deposited layers of a powder using a computer controlled laser that scans and sinters patterns onto the powder. SLS has the capability of processing a variety of polymer, metal, and ceramic powders directly into dense, functional forms. Commercial SLS technology only allows the processing of monolithic polymers, and of metal or ceramic powders with polymer binders to green bodies that can be postprocessed to near-full density. Feature resolution is limited to 250–500 μm at best and is defined by material, powder particle size (typically 10–100 μm), and focused laser beam diameter (typically 500 μm). With appropriate technological advances, a next-generation SLS system could incorporate the capability of constructing three-dimensional, geometrically and topologically complex, heterogeneous objects with features as small as 100 μm. This requires the design of a new material delivery system that can precisely deposit micro- to nanoscale powders in desired locations and the design of a sophisticated laser processing system for consolidating the deposited, patterned materials into desired shapes, and densities.

A new SLS system is currently being designed and developed by the first author (Das S). This system comprises a nozzle array-based deposition system that can deposit arbitrary patterns of multiple powder materials (0.1–100 μm particles) with a minimum in-plane 100 μm feature size and layer thicknesses ranging from 25 μm to 250 μm. A conceptual rendition of this system with powder reservoir illustrated for just one nozzle is shown in Fig. 4. Fortunately, all the aforementioned biomaterials are commercially available in powder form and are thus candidate materials for the new SLS system that will innovate the construction of bioimplants with functionally tailored geometry, composition, and porosity directly.

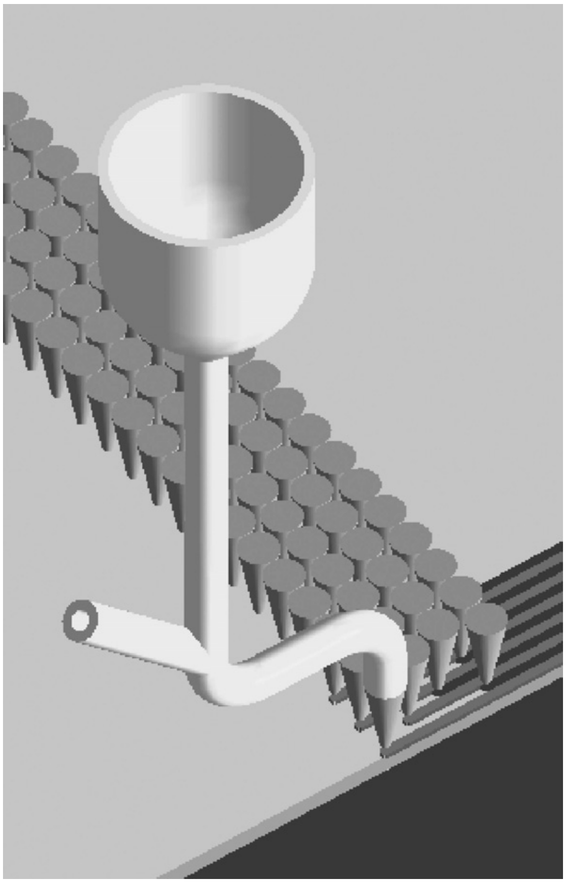
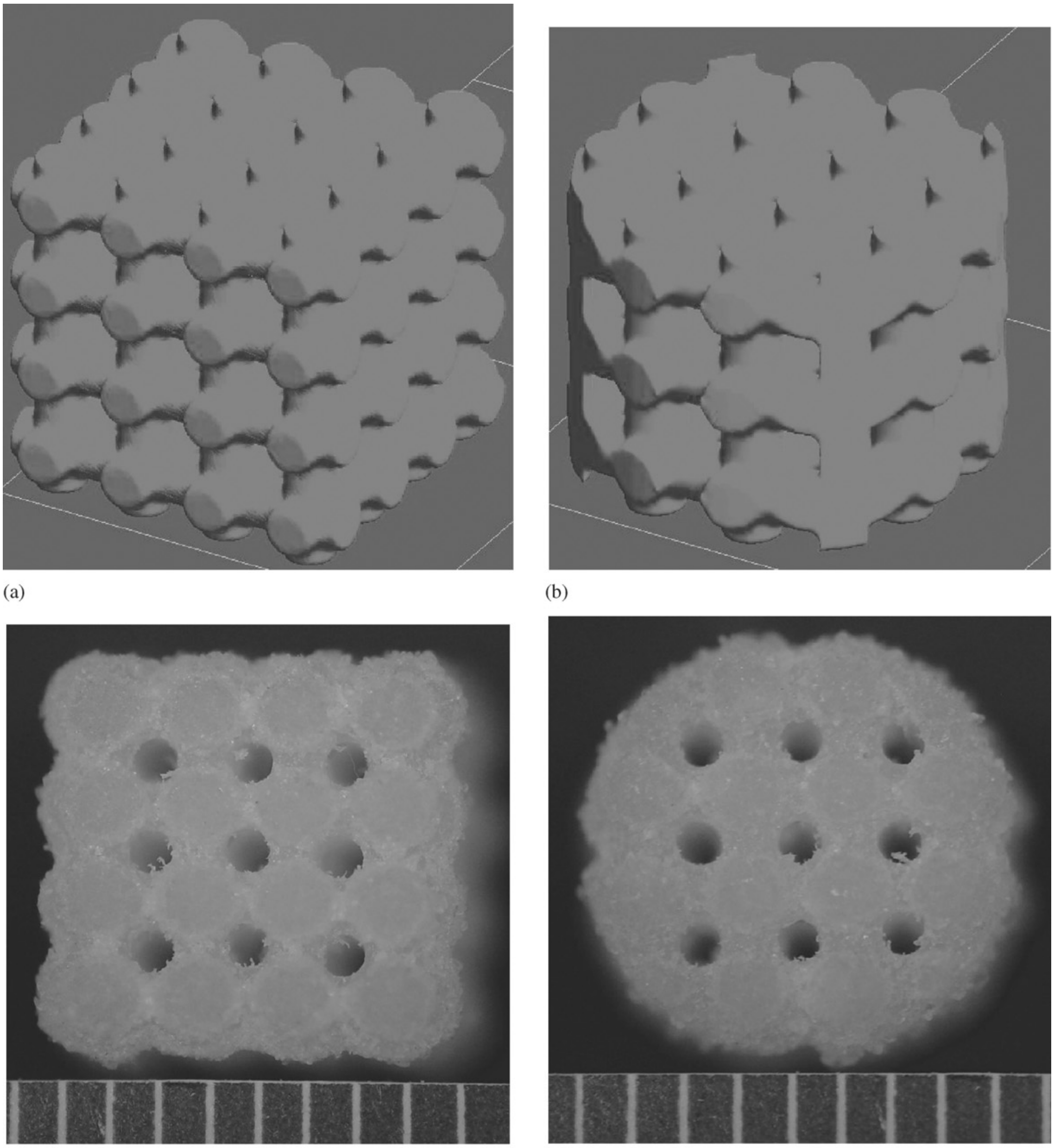


Figure 4
Rendition of nozzle array-based powder deposition system.

3. Recent Results and Future Challenges

Figure 5 shows examples of monolithic Nylon–6 scaffolds we have fabricated using SLS including cell-based designs with 800 μm orthogonal porosity and biomimetic scaffolds derived from human trabecular bone micro-CT data (scaled 4X).

Future challenges include fabricating monolithic scaffolds in PLA, PGA, PLGA copolymers and their composites with calcium phosphate ceramics as well as graded, heterogeneous scaffolds, first with discrete material interfaces and finally with smoothly varying compositions on length scales replicating interfaces in real tissue. Another significant challenge will be in evaluating the function and regenerative capability of the many scaffold materials and designs made possible by the integration of computational design, biomaterial SFF and synthesized biofactors. Scaffold function can be evaluated through standard tests, some of which fall under ASTM guidelines. Tissue regeneration presents a more difficult evaluation



(a) (b) (c) (d)

Figure 5
(a) 8 mm cubic cell-based scaffold design, (b) 8 mm diameter, 6 mm high cylindrical cell-based scaffold design (both with 800 μm orthogonal channels and 1200 μm pillars), (c) and (d) Scaffolds fabricated by SLS (scale is in mm)

challenge and while ASTM guidelines are currently being formulated (ASTM 2000), there remains no definite standard. Expense, time, and required surgical expertise will prohibit large animal tests of all the myriad materials and scaffold microstructures. Therefore, *in vitro* and small animal tests such as subcutaneous mouse models may present more tractable, yet incomplete tests for tissue engineering scaffolds. Perhaps the best approach will involve *in vitro* and small animal systems to screen scaffold materials and designs, taking only the most promising ones to large animal tests.

4. Summary

Together, computational design and SFF techniques are well-suited for creating tissue engineering scaffolds that replicate human anatomic external geometry with optimal internal architecture for tissue regeneration. Although much research on both SFF process development and biological testing remains to be done in the future, these techniques are expected to enable fabrication of functionally tailored scaffolds suitable for regenerating multiple tissues and tissue interfaces simultaneously.

Bibliography

ASTM 2000 Regenerative Medicine: Tissue Engineering Standards Come to ASTM, *ASTM Standardization News September*

Beaman J J, Barlow J W, Bourell D L, Marcus H L, Crawford R H, McAlea K P 1997 *Solid Freeform Fabrication: A New Direction in Manufacturing*. Kluwer

Bostrom R D, Mikos A G 1996 Tissue engineering of bone. In: Atala A, Mooney D J (eds.) *Synthetic Biodegradable Polymer Scaffolds*. Birkhauser, Boston Chap. 12

Bruder S P, Kraus K H, Goldberg V M, Kadiyala S 1998 Critical-sized canine segmental femoral defects are healed by autologous mesenchymal stem cell therapy. In: *Transactions of the 44th Annual Meeting of the Orthopaedic Research Society*, 147

Butler D L, Goldstein S A, Guilak F 2000 Functional tissue engineering: the role of biomechanics. *J. Biomech. Eng*. **122**, 570–5

Caplan A I, Bruder S P 1997 Cell and molecular engineering of bone regeneration. In: Lanza R P, Langer R, Chick W L (eds.) *Principles of Tissue Engineering*. Academic Press

Chu Gabriel T -M, *et al*. 1999 Ceramic SFF by direct and indirect stereolithography. *Mat. Res. Soc. Symp. Proc*. **542**, 119

Cima L G, Cima M J, 1996a *US Patent 5 490 962*

Cima L G, Cima M J, 1996b *US Patent 5 518 680*

Das S, *et al*. 2002 Computational Design, Freeform Fabrication and Testing of Nylon–6 Engineered Tissue Scaffolds, *Solid Freeform Fabrication Symposium Proceedings, 2002*

Giordano R A, *et al*. 1996 Mechanical properties of dense polylactic acid structures fabricated by three dimensional printing. *J. Biomater. Sci. Polym. Ed*. **8** (1), 63–75

Hollister S J, Kikuchi N 1994 Homogenization theory and digital imaging: a basis for studying the mechanics and design principles of bone tissue. *Biotechnol. Bioeng*. **43**, 586–96

Hollister S J, Chu T M, Guldberg R E, Zysset P K, Levy R A, Halloran J W, Feinberg S E 1999 Image based design and manufacture of scaffolds for bone reconstruction. In: Pedersen P, Bendsoe M P (eds.) *Synthesis in Bio Solid Mechanics*. Kluwer, 163

Hollister S J, Maddox R D, Taboas J M Optimal design and fabrication of scaffolds to mimic tissue properties and satisfy biological constraints. *Biomaterials* **23**, 4095–4103

Krebsbach P H, Kuznetsov S A, Satomura K, Emmons R V B, Rowe D W, Robey P G 1997 Bone formation in vivo: comparison of osteogenesis by transplanted mouse and human marrow stromal fibroblasts. *Transplantation* **63**, 1059

Kuboki Y, Saito T, Murata M, Takita H, Mizuno M, Inoue M, Nagai N, Poole A R 1995 Two distinctive BMP-Carriers induce zonal chondrogenesis and membranous ossification, respectively; geometrical factors of matrices for cell-differentiation. *Connective Tissue Res*. **32**, 219

Kuboki Y, Takita H, Kobayashi D, Tsuruga E, Inoue M, Murata M, Nagai N, Dohi Y, Ohgushi H 1998 BMP-induces osteogenesis on the surface of hydroxyapatite with geometrically feasible and nonfeasible structures: topology of osteogenesis. *J. Biomed Mat. Res*. **39**, 190

LeBaron R G, Athanasiou K A 2000 *Ex vivo* synthesis of articular cartilage. *Biomaterials* **21**, 2575–87

Putney S D 1999 Advances in protein drug delivery. *Pharmaceutical News* **6** (2)

Reddi A H 1998 Initiation of fracture repair by bone morphogenetic proteins. *Clin. Orthop*. **355**, S66

Risbud M V, Bhonde R R 2001 Polyamide 6 composite membranes: properties and *in vitro* biocompatibility evaluation. *J. Biomater. Sci. Polymer Edn*. **12** (1), 125–36

Rothen–Weinhold A, *et al*. 1999 Injection–molding versus extrusion as manufacturing technique for the preparation of biodegradable implants. *Euro. J. Pharm. Biopharm*. **48**, 113–21

Sampath T K, Reddi A H 1984 Importance of geometry of the extracellular matrix in endochondral bone differentiation. *J. Cell Biol*. **98**, 2192

Stamboulis A G, *et al*. 2002 Novel biodegradable polymer/bioactive glass composites for tissue engineering applications. *Adv. Eng. Mater*. **4** (3), 105–9

Stancari F, *et al*. 2000 Use of PLA–PGA (Copolymerized Polylactic/Polyglycolic acids) as a bone filler: clinical experience and histologic study of a case. *Quintessenz (Germany)* **51** (1), 47–52

Taboas J M, Maddox R D, Krebsbach P H, Hollister S J Indirect solid free form fabrication of local and global porous, biomimetic and composite 3D polymer-ceramic scaffolds. *Biomaterials* **24**, 181–94

Tsuruga E, Takita H, Itoh H, Wakisaka Y, Kuboki Y 1997 Pore size of porous hydroxyapatite as the cell-substratum controls BMP-induced osteogenesis. *J. Biochemistry* **121**

Vacanti J P, Cima L G, Cima M J 2001 *US Patent 6 176 874*

Weiss L E, Calvert J W 2000 *US Patent 6 143 293*

Yang S, Leong K F, Du Z, Chua C K 2001 The design of scaffolds for use in tissue engineering: Part 1. Traditional factors. *Tissue Eng*. **7**, 679–89

Yazemski M J, Payne R G, Hayes W C, Langer R S, Mikos A G 1996 Evolution of bone transplantation: molecular, cellular, and tissue strategies to engineer human bone. *Biomaterials* **17**, 175

Zein I, *et al*. 2002 Fused deposition modeling of novel scaffold architectures for tissue engineering applications. *Biomaterials* **23** (2), 1169–85

S. Das and S. J. Hollister
University of Michigan, USA

Titanium Alloys: Alloy Designation System

Titanium alloys can be divided into four categories; the near alpha alloys, the alpha + beta alloys, the beta

Table 1
Table of titanium alloy designations.

Alloy composition	UNS number	ASTM designation	Comments
Unalloyed titanium	R50250	Grade 1	Grades 2-4 have increased strength due to higher oxygen contents BT1-0 (purer) and BT1-0 Russian
Ti-0.2Pd	R52400 and R52250	Grade 7 and 11	Corrosion resistant Ru substitution reduces cost
Ti-0.3Mo-0.8Ni	R53400	Grade 12	Corrosion resistant
Ti-3Al-2.5V	R56320	Grade 9	Formable, tubing
Ti-5Al-2.5Sn	R54520		Weldable, cryogenic use Russian BT5-1
Ti-6Al-7Nb			Timetal 367, excellent biocompatibility
Ti-6Al-2Sn-4Zr-2Mo-0.1Si	R54620		Creep resistant
Ti-8Al-1Mo-1V	R54810		High modulus
Ti-6Al-2.7Sn-4Zr-0.4Mo-0.45Si			Timetal 1100, use to 600 °C (1100 °F)
Ti-2.5Cu			IMI-230[a]
Ti-5Al-3.5Sn-0.3Zr-1Nb-0.3Si			IMI-829[a] Creep resistant
Ti-6Al-4Sn-4Zr-1Nb-0.5Mo-0.4Si			IMI-834[a] Creep resistant
Ti-4.3Al-1.4Mn			Russian OT4 structural
Ti-6.7Al-3.3Mo-0.3Si			Russian BT8 high temperature
Ti-6.4Al-3.3Mo-1.4Zr-0.28Si			Russian BT9 high temperature
Ti-7.7Al-0.6Mo-11Zr-1.0Nb-0.12Si			Russian BT18 high temperature
Ti-6Al-4V	R56400	Grade 5	Work-horse alloy Russian BT6
Ti-6Al-4V ELI	R56401		Low interstitial, damage tolerant
Ti-6Al-6V-2Sn	R56620		Higher strength than Ti-6Al-4V
Ti-4Al-4Mo-2Sn-0.5Si			Timetal 500
Ti-5.5Al-5V-5Mo-3Cr			Timetal 555, High strength forgings
Ti-5Al-1Sn-1Zr-N-0.8Mo			Timetal 5111 High toughness, heavy section weldable
Ti-4Al-4Mo-4Sn-0.5Si			Timetal 551

Table 1
Continued

Alloy composition	UNS number	ASTM designation	Comments
Ti-4.5Al-3V-2Mo-2Fe			Superplastic alloy SP-700
Ti-6Al-1.7Fe-0.1Si			Timetal 62S low cost alloy
Ti-5Al-2Sn-2Zr-4Mo-4Cr	R58650		Ti-17, high strength, moderate temperature
Ti-4.5Al-5Mo-1.5Cr			CORONA 5, high fracture toughness
Ti-6Al-2Sn-4Zr-6Mo	R56260		Moderate temperature, strength and long-term creep
Ti-6Al-2Sn-2Zr-2Mo-2Cr-0.25Si			Strength-toughness
Ti-3Al-8V-6Cr-4Mo-4Zr	R58640		Beta C (38-6-44)
Ti-10V-2Fe-3Al			Ti-10-2-3, high strength forgings
Ti-15V-3Al-3Cr-3Sn			Ti-15-3, high strength and strip processable
Ti-3Al-7.4Mo-10.5Cr			Russian BT15 structural
Ti-1.5Al-5.5Fe-6.8Mo			Timetal LCB low cost beta
Ti-15Mo-3Al-2.7Nb-0.25Si	R58210		Timetal 21S
Alpha-2 (Ti3Al) aluminide			Experimental intermetallics
Gamma (TiAl) aluminide			Two-phase alloys ($\alpha_2+\gamma$) look best, semi-commercial
Ti-Ni			Shape memory alloy
CermeTi			Particle reinforced, with TiC, TiB_2 or TiAl

[a] IMI—Imperial Metals Industries (now part of Timet).

alloys, and the titanium aluminide intermetallics. The first three categories relate to the predominant phase present: the low-temperature (hexagonal close packed) alpha or the elevated-temperature (body-centered cubic) beta. The intermetallics (Ti_xAl, where $x=1$ or 3) exhibit an ordered structure. The major characteristics of all four classes of alloys are summarized in *Titanium: Alloying*.

The titanium industry has not used a broad-based alloy designation system, as with aluminum alloys. However, in recent years a UNS system has been introduced to complement a (partial) ASTM system and some company/country designations. Some commonly used alloys are listed in Table 1; further information can be found in the cited references.

Bibliography

Boyer R, Welsch G, Collins E W (eds.) 1994 *Materials Properties Handbook, Titanium Alloys.* ASM, Materials Park, OH

Titanium Metals Corporation 1996 *Titanium Alloys.* Timet, Denver, CO

F. H. Froes
University of Idaho, Moscow, Idaho, USA

Titanium: Alloying

1. Alloying Behavior of Titanium

Titanium exists in two crystalline states: a low-temperature alpha (α) phase that has a close-packed hexagonal (c.p.h.) crystal structure, and a high-temperature beta (β) phase that has a b.c.c. structure (Fig. 1). This allotropic transformation (or beta transus temperature) occurs at 880 °C (1620 °F) in nominally pure titanium. This temperature is the key to the processing of titanium alloys. Titanium has a number of features, noted below, which make it very different from other light metals such as aluminum and magnesium. Exhibiting an allotropic transformation allows for the formation of alloys composed of α-, β-, or α/β-microstructures in addition to compound formation. Alloying elements in titanium can also be divided into substitutional and interstitial solutes, i.e., those elements that substitute for titanium on lattice sites (molybdenum, vanadium, manganese, etc.) and those that "squeeze in" (oxygen, nitrogen, hydrogen) between the titanium atoms, respectively. Because of its electronic structure, titanium can form

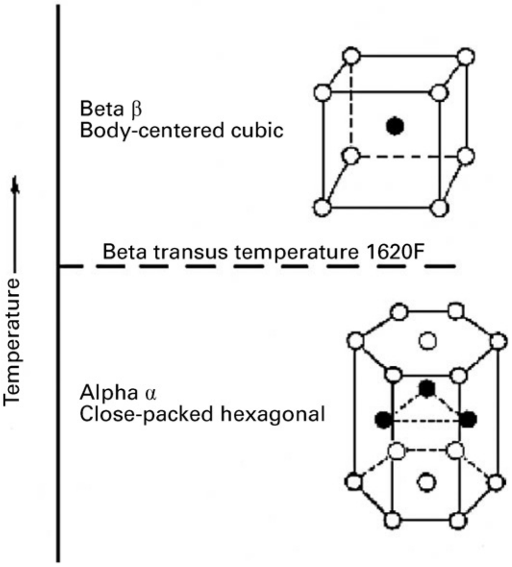

Beta β
Body-centered cubic
Beta transus temperature 1620F
Temperature
Alpha α
Close-packed hexagonal

Figure 1
Crystal structures of titanium.

solid solutions with most substitutional elements having a size factor within 20%, giving the opportunity for many alloying possibilities. Titanium also reacts strongly with interstitial elements such as nitrogen, oxygen, and hydrogen at temperatures below its melting point. When reacting with other elements, titanium may form solid solutions and compounds with metallic, covalent, or ionic bonding.

The choice of alloying elements is determined by the ability of the element to stabilize either the α- or β-phases (Fig. 2), i.e., raising or lowering the beta transus temperature, respectively. Aluminum, oxygen, nitrogen, and carbon are the most common α-stabilizing elements (see below). Zirconium, tin, and silicon are viewed as neutral in their ability to stabilize either phase. Elements that stabilize the β-phase can either form binary systems of the β-isomorphous type or of the β-eutectoid type (see below). Elements forming isomorphous-type binary systems include molybdenum, vanadium, and tantalum, while copper, manganese, chromium, iron, nickel, cobalt, and hydrogen are eutectoid formers. The β-isomorphous alloying elements do not form intermetallic compounds while the β-eutectoid formers do. The former type of β-stabilizing additions have traditionally been preferred to the eutectoid-type elements as additional to α/β- or β-alloys to improve hardenability and increase response to heat treatment.

There have been a number of attempts to establish methods for predicting the constitution of titanium alloys. While each has achieved limited success, none has been able to deal successfully with a wide range of alloy compositions or a wide range of alloying

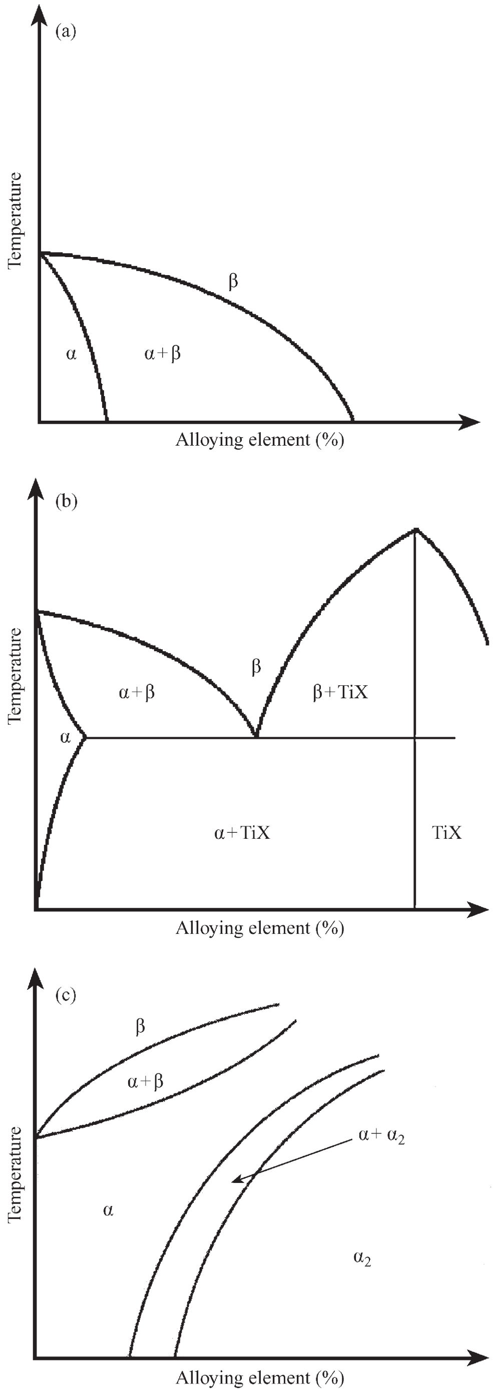

(a)
Temperature
β
α
α+β
Alloying element (%)
(b)
Temperature
β
α+β
β+TiX
α
α+TiX
TiX
Alloying element (%)
(c)
β
α+β
α+ α2
Temperature
α
α2
Alloying element (%)

Figure 2
Schematic phase diagrams showing the effect of alloying elements in stabilizing either the α- or β-phases. (a) β-Isomorphous; (b) β-eutectoid formers; (c) α-stabilizing system in which a DO_{19} (α_2) compound forms at higher alloying (e.g., aluminum) contents.

element types. Thus, the constitution of binary titanium alloys must still be determined experimentally. Such experimental determinations are often complicated by the presence of oxygen, hydrogen, or other impurities. In such cases the phase boundaries that are experimentally determined may be subject to significant error.

A number of authors have attempted to categorize titanium alloy phase diagrams. All agree that there are two major divisions: α- and β-stabilized systems. Of these, perhaps the most convenient is that developed by Molchanova (Fig. 3). Here the α-stabilizers are divided into those having complete stability, in which the α-phase can coexist with the liquid (e.g., Ti–O and Ti–N) and there is a simple peritectic reaction, and those that have limited α-stability in which, with decreasing temperature, decomposition of the α-phase takes place by a peritectoid reaction into β-phase plus a compound (β-peritectoid).

Examples of the latter type of system are Ti–B, Ti–C, and Ti–Al. Molchanova also divides the β-stabilizers into two categories: β-isomorphous and β-eutectoid. In the former system an extreme β-solubility range exists with only a restricted α-solubility range. Examples are Ti–Mo, Ti–Ta, and Ti–V, with elements such as zirconium and hafnium occupying an intermediate position since they have complete mutual solubility in both the α- and β-phases. For the β-eutectoid systems the β-phase has a limited solubility range and decomposes into α-phase and a compound (e.g., Ti–Cr and Ti–Cu). This category can also be further subdivided depending on whether the β-transformation is rapid (the ‘‘active’’ eutectoid formers such as Ti–Si, Ti–Cu, and Ti–Ni) or slow (the ‘‘sluggish’’ eutectoid formers such as Ti–Cr and Ti–Fe).

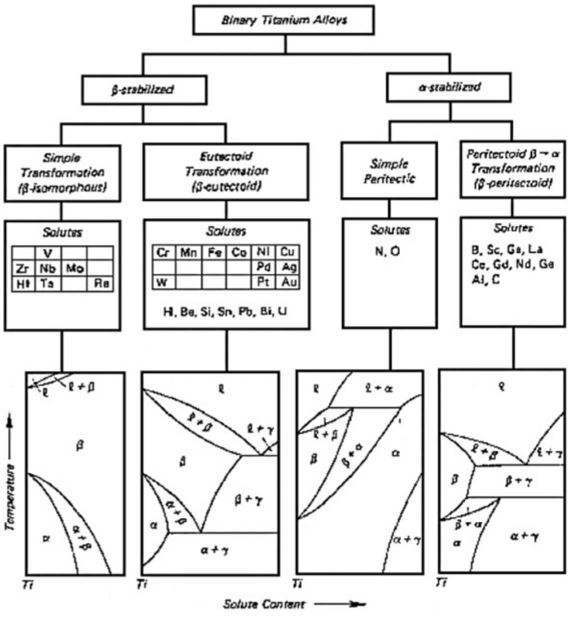


Figure 3
Categorization of titanium phase diagrams.

Titanium alloys are categorized into one of four groups: α-, α/β-, β-alloys, and the intermetallics (Ti_xAl, where $x = 1$ or 3). Titanium alloys for aerospace application contain α- and β-stabilizing elements to achieve the required mechanical properties such as tensile strength, creep, fatigue, fatigue crack propagation resistance, fracture toughness, stress-corrosion cracking, and resistance to oxidation. Once the chemistry is selected, optimization of mechanical properties is achieved by deformation/working to control the size, shape, and dispersion of both the β- and α-phases. For nonaerospace applications, in which titanium is used because of its excellent corrosion resistance, the commercially pure grades and alloys with addition such as the platinum group metals (PGM) are used.

1.1 α-Alloys

The α-alloys contain predominantly α-phase at temperatures well above 540 °C (1000 °F). A major class of α-alloy is the unalloyed commercially pure (CP) titanium family of alloys that differ by the amount of oxygen and iron in each alloy. Alloys with higher interstitial content are higher in strength, hardness, and transformation temperature compared to high-purity alloys. Other α-alloys contain additions such as aluminum and tin (e.g., Ti–5Al–2.5Sn and Ti–6Al–2Sn–4Zr–2Mo (in wt.%)). Generally, α-rich alloys are more resistant to high-temperature creep than α/β- or β-alloys, and α-alloys exhibit little strengthening from heat treatment. These alloys are usually annealed or recrystallized to remove stresses from cold working. They have good weldability and generally inferior forgeability in comparison to α/β- or β-alloys.

1.2 α/β-Alloys

α/β-Alloys contain one or more of the α- and β-stabilizers. These alloys retain more β-phase after final heat treatment than the near α-alloys, and can be strengthened by solution treating and aging, although they are generally used in the annealed condition. Solution treatment is usually performed high in the α/β-phase field followed by aging at lower temperature to precipitate α-phase, giving a mixture of relatively coarse primary α- and fine α-phase in an α/β-matrix. Solution treating and aging can increase the strength of these alloys by up to 80%. Alloys with low amounts of β-stabilizer (e.g., ‘‘work-horse’’ Ti–6Al–4V (wt. %), alloy) have poor hardenability and must be rapidly quenched for subsequent strengthening. A water quench of Ti–6Al–4V will adequately harden sections only less than 25 mm (1 in). β-Stabilizers in α/β-alloys increase hardenability.

1.3 β-Alloys

β-Alloys have more β-stabilizer and less α-stabilizer than α/β-alloys. These alloys have high hardenability with the β-phase being retained completely during air cooling of thin sections and water quenching of thick sections. B-Alloys have good forgeability and good cold formability in the solution-treated condition. After solution treatment, aging is performed to transform some β-phase to α-phase. The strength level of these alloys can be greater than that of α/β-alloys, a result of finely dispersed α-particles in the β-phase. These alloys have relatively higher densities, and generally lower creep strengths than the α/β-alloys. The fracture toughness of aged β-alloys at a given strength level is generally higher than that of an aged α/β-alloy, although crack growth rates can be faster.

1.4 Titanium Aluminides

To increase the efficiency of gas turbine engines, higher operating temperatures are necessary, requiring low-density alloys with enhanced mechanical properties at elevated temperatures. The titanium aluminide intermetallic compounds Ti_3Al (α_2) and TiAl (γ) show potential for applications to as high as 900 °C (1650 °F). The major disadvantage of this alloy group is low ambient temperature ductility. However, it has been found that niobium, or niobium with other β-stabilizing elements, in combination with microstructure control, can significantly increase room-temperature ductility in the Ti_3Al alloys. By careful control of the microstructure, the ambient temperature ductility of two-phase TiAl ($\gamma+\alpha_2$) has been increased to levels as high as 5% elongation. The TiAl compositions (e.g., Ti–48Al–2Cr–2Nb (at.%)) have now reached a stage of maturity where they are serious contenders for use in advanced gas turbine engines and automobiles.

2. Summary

The type and concentration of alloying elements affects the equilibrium constitution of titanium alloys by preferentially stabilizing one or the other of the two allotropic forms. In addition, the alloying elements partition selectively to one or the other constituent and provide solid-solution strengthening. Aside from its effects on equilibrium constitution, alloying also affects the kinetics of decomposition of the elevated-temperature β-phase. The resulting microstructures have a strong effect on mechanical properties, so alloying also affects the properties of titanium alloys by influencing the evolution of their microstructure.

Bibliography

Blenkinsop P A, Evans W J, Flowers H M 1995 *Titanium '95 Science and Technology*. Institute of Materials, London

Froes F H, Caplan I L 1993 *Titanium '92 Science and Technology*. TMS, Warrendale, PA

Froes F H, Eylon D, Bomberger H B (eds.) 1985 *Titanium Technology: Present Status and Future Trends*. Titanium Development Association (now International Titanium Association), Boulder, CO

Molchanova E K 1965 *Phase Diagrams of Titanium Alloys* [Transl. of *Atlas Diagram Sostoyaniya Titanovyk Splavov*]. Israel Program for Scientific Translations, Jerusalem

Williams J C, Starke E A 1984 *The Role of Thermochemical Processing in Tailoring the Properties of Aluminum and Titanium Alloys*, ASM Seminar Series. American Society for Metals, Metals Park, OH

F. H. Froes
University of Idaho, Moscow, Idaho, USA

Tool and Die Steels

Any steel that is used for the working parts of tools may be called a tool steel. If plain carbon steel had been satisfactory for all tooling problems, there would have been no need for alloy steels. Alloying elements are added to plain carbon steel to enable it to attain properties that otherwise it could not. These properties can be listed under four headings:

(i) greater wear resistance for cutting or abrasion;
(ii) greater toughness or strength;
(iii) hardening accuracy and safety—and increased hardenability; and
(iv) "red hardness," or the ability of the tool to do its work when it is heated so hot that a plain carbon steel would soften.

This article is an adaptation of the classic monograph of Roberts and Cary (1980). The effects of alloying elements on tool steel properties and structure are discussed in detail, then the types, properties, and selection of tool and die steels are briefly discussed.

1. Alloying Element Effects

Steels used for tools generally contain at least 0.60% carbon in order to assure the attainment of a martensitic hardness of at least 60 HRC. Carbon in excess of this minimum is employed only to have undissolved carbides in the martensitic structure to increase the resistance to wear (Brick *et al.* 1965). Alloying elements modify the properties of tool steels by:

(i) changing the composition and properties of the carbide phase;
(ii) changing the hardening characteristics, including: critical temperature and excess carbide solution on heating, grain growth during heating,

hardenability and transformation products on cooling, martensite formation and austenite retention on cooling (Ms and Mf temperatures); and

(iii) changing the tempering characteristics on reheating.

The alloying element effects listed above will be discussed in detail in this article.

1.1 Stability of Carbides

The alloying elements and their carbide types can be arranged in the following order of decreasing carbide stability: titanium (TiC), tungsten (M_6C, W_2C, $M_{23}C_6$), molybdenum (M_6C, W_2C, $M_{23}C_6$), vanadium (VC), chromium (Cr_7C_3, $M_{23}C_6$, M_3C), manganese (M_3C), iron (Fe_3C), cobalt, nickel, and silicon. Proceeding from the beginning to the end of the list, the alloys exhibit the following differences in characteristics. Increasing tendency to combine preferentially with carbon and take it away from the elements lower in the list in a multiple alloy steel. Increasing hardness and wear resistance of the resulting carbides. Increasing reluctance of these carbides to dissolve in the matrix on heating for hardening, necessitating higher temperatures or longer times for solution to occur. This results in sharply improved resistance to grain growth and an increasing proportion of the harder, more wear resistant carbides in the excess carbide structure of the hardened steel.

1.2 Critical Temperature and Excess Carbides

Alloying elements can effect the following changes in the heating portion of the hardening cycle: (i) raise or lower the critical temperature for austenite formation; (ii) change the quantity of undissolved or excess carbide in the hardened structure; (iii) Modify the rate of solution of the excess carbide; and (iv) alter the grain coarsening characteristics.

1.3 Grouping of Alloying Elements with Respect to their Effect on Hardening Temperature Range

The following ferrite formers restrict the austenite range and consequently raise the hardening temperature: titanium, vanadium, molybdenum, silicon, tungsten, and chromium. The following austenite formers expand the austenite range and consequently lower the hardening temperature: carbon, nitrogen, nickel, manganese, copper, and cobalt. (The two lists are arranged roughly in order of decreasing potency.)

1.4 Effect of Alloying Elements on Grain Size

In view of their close relationship to carbide reactions, the alloying elements can be expected to have a profound effect on grain size. This effect is actually the net result of several factors. (i) Effect of alloy on hardening range. Higher heating temperatures provide greater driving force for grain growth, but may be necessary for full hardening. (ii) Effect of alloy on properties of excess carbide, principally its temperature of rapid solution. For grain refinement, some excess carbide should remain undissolved at the normal hardening temperature. (iii) Effect of alloy on quantity of excess carbide. More carbides give a greater opportunity for keying and grain refinement. The effectiveness of factors (ii) and (iii) in opposing factor (i) determines the ability of the alloy to refine the grain size. In general, the alloying elements restrict grain growth in direct proportion to their carbide-stabilizing ability.

1.5 Hardenability and Transformation Products

Alloy additions to tool steels affect the cooling portion of the hardening cycle in the following ways.

(i) Move the isothermal transformation curve to longer or shorter times, thereby changing the cooling rate and quenching medium required for full hardening.

(ii) Alter the shape of the isothermal transformation curve, thereby changing the structures obtained and the temperature ranges of transformation at cooling rates below those required for full hardening. (The carbide-forming elements chromium, molybdenum, tungsten, and vanadium tend to alter the shape of the isothermal transformation curve in addition to shifting it to longer times.)

(iii) Shift the Ms and Mf points for martensite transformation to higher or lower temperatures, thereby changing the amount of austenite retained on quenching or after refrigeration treatment. All elements except cobalt and aluminum decrease the martensite start temperature and increase the amount of retained austenite. Carbon is the most potent of the alloying elements for lowering the Ms temperature and increasing retained austenite.

(iv) Change the susceptibility of the steel to austenite stabilization from slow cooling or isothermal holding at temperatures below Ms, thereby further altering the amount of retained austenite after quenching.

With the exception of cobalt, all of the alloying elements added to tool steels (including carbon) tend to move the isothermal transformation curve to longer times for greater hardenability with less severe quenching media.

There is, however, a major factor unique to tool steels that must always be kept in mind—*the effect of undissolved carbide*. Strong carbide formers like vanadium or tungsten can tie up carbon so effectively that it resists solution in the austenite during heating for hardening. In such cases, the austenite may be lower in carbon content than the corresponding unalloyed steel and the isothermal transformation curve is shifted to shorter rather than longer times. Thus, when considering the effect of alloys on tool steel isothermal transformation curves, the hardening temperature and the presence or absence of undissolved carbides must always be kept in mind.

1.6 Effect of Alloying Elements on Tempering Characteristics

Reheating for tempering is an essential final step in the heat treatment of tool steel. It provides for the precipitation of carbides dissolved during hardening into a fine dispersion for increased toughness and the desired working hardness.

Secondary hardening of production tool steel is produced by a combination of two reactions: (i) transformation of retained austenite; and (ii) precipitation of submicroscopic alloy carbides in the tempered martensite.

All steels in the untempered condition contain a percentage of retained austenite. In low- to medium-alloy steels, this austenite transforms to bainite or is partially stabilized at the low tempering temperatures used. In high-alloy secondary hardening steels, the retained austenite remains untransformed up to 427–538 °C (800–1000 °F) where a conditioning reaction occurs. The austenite transforms to martensite on subsequent cooling from the tempering temperature. A second tempering treatment is then needed to temper this newly formed martensite and toughen the structure.

The second reaction, precipitation of submicroscopic alloy carbides, also occurs in the temperature range 482–538 °C (900–1000 °F). Research indicates that only three alloy carbides are capable of producing this effect: W_2C, Mo_2C, or VC. Sufficient amounts of the elements tungsten, molybdenum, or vanadium must be present to precipitate these carbides in quantity and obtain the rehardening effect. Note that chromium additions sharply retard the rate of softening during tempering and produce no true secondary hardening.

It should be remembered that the effects of these carbide-forming elements on tempering depend, as in previously discussed reactions, on prior heating for hardening. The effects described are for carbon and alloys completely dissolved in the austenite. Undissolved carbides reduce the effectiveness in proportion to the carbon and alloys they keep from solution. Thus, in addition to the retained austenite effect, the hardening treatment also can completely change the tempering curve of carbide-containing steels by altering the amount of dissolved carbon and alloy available for precipitation on tempering.

Of the remaining elements, silicon retards softening in a similar manner to chromium but to less elevated temperatures, while manganese and nickel are relatively ineffective for tempering resistance. Cobalt enhances the effect of tungsten and/or molybdenum by moving secondary hardening and hot hardness retention to slightly higher temperatures, and is included in high-speed steels for that purpose.

2. Types, Properties, and Selection of Tool and Die Steels

When the properties required for specific applications are analyzed, it is found that the applications for tool and die steels generally require balancing of the four properties discussed in the introduction to this article:

- wear resistance;
- toughness or strength;
- hardening accuracy, hardening safety, and increased hardenability; and
- "red hardness."

Over the years many different tool and die steels have been developed that balance these properties for specific applications. Other than the high-speed steels, tool and die steels can be divided into five major classifications: carbon tool steels, low-alloy tool steels, special purpose tool steels, cold work die steels, and hot work die steels. Each of these major groups can be further divided into classes depending on their alloying elements, heat treatment, or properties. Space does not permit a detailed discussion of the tool steels in these classes or how to select tool steels for specific applications. The interested reader is referred to Roberts and Cary (1980) for more information.

However, as an introduction to properties and selection, the remainder of this article briefly discusses 12 of the most popular tool and die steels and some of their properties. This discussion is based on a useful selector chart, shown in Fig. 1, which has been developed to enable toolmakers to choose the proper tool steels from 12 of the most popular grades to obtain the balance of properties that a particular application requires (Palmer *et al.* 1978). It must be emphasized that this chart should only be used as a starting point for tool steel selection. As technology has advanced, many other tool and die steels have been developed to enable toolmakers to further improve their tools. A tool steel metallurgist should be consulted before making any tool and die steel selection for a particular application.

In the following discussion the class of tool steel is listed first followed by the AISI designation for the most popular grade in the class. Its nominal composition (in wt.%) and how it is quenched from the hardening temperature are listed in parentheses

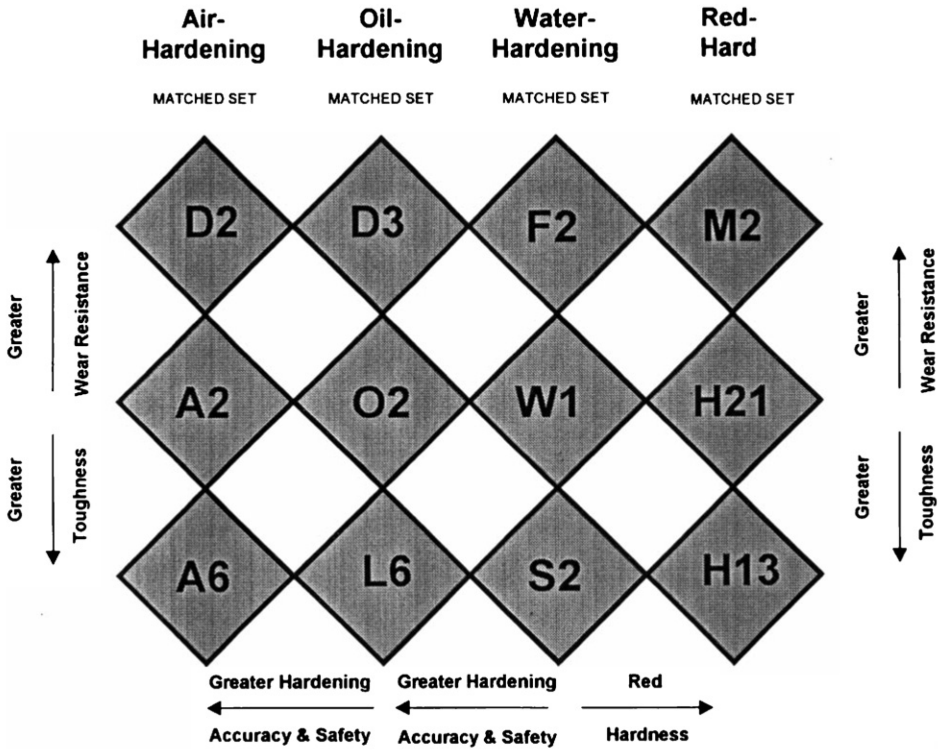

Air-Hardening
Oil-Hardening
Water-Hardening
Red-Hard
MATCHED SET
MATCHED SET
MATCHED SET
MATCHED SET
D2
D3
F2
M2
A2
O2
W1
H21
A6
L6
S2
H13
Greater Wear Resistance
Greater Toughness
Greater Wear Resistance
Greater Toughness
Greater Hardening Accuracy & Safety
Greater Hardening Accuracy & Safety
Red Hardness

Figure 1
Selector chart using the matched set diagram as a tool steel guide. (i) Use an *air-hardening steel* if the tool requires extreme accuracy and safety in hardening—or if deep hardening is required in large sizes: D2, for maximum wear resistance with fair toughness, plus extreme hardening accuracy and safety; A2, for good wear resistance and good toughness, plus extreme hardening accuracy and safety; A6, for maximum toughness and fair wear resistance, plus extreme hardening accuracy and safety. In case of doubt, always start with the *tougher* of the two steels in question. (ii) Use an *oil-hardening steel* if the tool is of intricate shape that may crack if quenched in water—or if the tool must hold size and shape accurately during hardening: D3, for maximum wear resistance with fair toughness, plus hardening accuracy and safety; O2, for good wear resistance and good toughness, plus hardening accuracy and safety; L6, for maximum toughness and fair wear resistance, plus hardening accuracy and safety. In case of doubt, always start with the *tougher* of the two steels in question. (iii) Use a *water-hardening steel* if there is little danger of hardening cracks, and if size change and warpage are not involved: F2, for maximum wear resistance with fair toughness; W1, for good wear resistance and good toughness; S2, for maximum toughness and fair wear resistance. In case of doubt, always start with the *tougher* of the two steels in question. (iv) Use a *red-hard steel* only for tools that become heated higher than about 300 °F (150 °C) in service—otherwise use the water-, oil-, or air-hardening matched set: M2, a high-speed steel for metal cutting tools requiring red hardness, keen cutting qualities, and fine-grained toughness; H21, for hot work dies and tools requiring good wear resistance and good toughness; H13, for hot work dies and tools requiring maximum toughness and fair wear resistance. (The diamond diagram is a registered trademark of CRS Holdings, Inc.)

prior to a discussion of the properties and applications for the particular tool and die steel.

2.1 Carbon Tool Steels

AISI W1 (0.6–1.4 C, 0.25 Mo, 0.25 Si; water hardening).

Plain carbon steels are among the least expensive of tool steels, and their properties provide a convenient baseline against which all other tool steels may be compared. Since the wear resistance of plain carbon tool steel is lower than that of alloy steels of equivalent carbon content, the useful life of a plain carbon tool will normally be shorter than that of a tool made from a higher-alloy steel. One notable exception is found in cold header dies where, owing to their combination of hard case and tough core, plain carbon tools steels are unsurpassed in performance.

2.2 *Silicon Tool Steels*

AISI S2 (0.5 C, 0.5 Mn, 1.0 Si, 0.2 V, 0.5 Mo; water hardening).

Special purpose tool steels in this class, originally used for springs, are useful primarily for their shock resistance, high fatigue resistance, and good wear ability. They can be treated to a fairly high hardness, making them applicable to many punch and die uses as well as for shock-resisting tools. Silicon tool steels may be differentiated from most other classes of tool steels on the basis of their high toughness.

2.3 *Tungsten Finishing Steels*

AISI F2 (1.25 C, 0.25 Mn, 0.25 Si, 3.5 W, 0.3 Mo; water hardening).

Special purpose tungsten finishing steels are generally used where extreme wear resistance and the ability to retain a keen edge are desired. Typical applications include tube and wire drawing dies and roll turning tools. Under certain conditions, wear resistance of the tungsten finishing tools is 4–10 times that of plain carbon steel at maximum hardness.

2.4 *Oil-hardening Cold Work Die Steels*

AISI O2 (0.95 C, 1.6 Mn, 0.25 Si, 0.2 Cr, 0.2 V, 0.5 W; oil hardening).

Cold work die steels are among the most important in the tool steel classification because they can be used for so many types of tool and die work. Oil-hardening types are probably more widely used than any other types of tool steels, except carbon and high-speed steels. Their properties include:

(i) low movement in hardening;
(ii) high as-quenched hardness;
(iii) high hardenability from low quenching temperatures;
(iv) relative freedom from cracking on quenching intricate sections; and
(v) maintenance of a keen edge for cutting purposes.

However, they possess no "red hardness" properties that would permit cutting at high speeds or that would allow their use for hot working. Typical applications include taps, reamers, broaches, blanking dies, forming dies, gauges, form tools, hobs, thread rolling dies, master tools, punches, burnishing tools, knurling tools, feed rolls, small shear blades, slitting saws, circular cutters, drills, coining dies, cold trimming dies, plastic molds, and drawing dies.

2.5 *Low-alloy Tool Steels (Based on Nickel and Greater than* 0.65% *Carbon)*

AISI L6 (0.7 C, 0.55 Mn, 0.25 Si, 0.75 Cr, 1.5 Ni, 0.3 Mo; oil hardening).

The steels in this class may be thought of as plain carbon steels to which small amounts of nickel or chromium, or both, plus manganese and molybdenum are added to obtain increased hardenability. Nickel, however, an austenite stabilizer and noncarbide former, behaves somewhat differently from chromium, which is a ferrite former and a relatively strong carbide former. Typical applications include woodworking saws and knives, shear blades, blanking dies, punches, press-brake dies, and such nontool uses as spindles, clutch parts, gears, and ratchets.

2.6 *High-carbon, High-chromium Cold Work Die Steels*

AISI D2 (1.5 C, 0.3 Mn, 0.25 Si, 12 Cr, 0.6 V, 0.8 Mo; air hardening); AISI D3 (2.2 C, 0.3 Mo, 0.25 Si, 12 Cr, 0.5 Ni, 0.6 V; oil hardening).

High-carbon, high-chromium steels, first developed commercially as a possible substitute for high-speed steel cutting tools during World War I, were used only to a limited extent for cutting tools because they had insufficient hot hardness and proved to be too brittle for this purpose. However, it was recognized early on that high wear resistance imparted by numerous hard chromium carbides combined with remarkable nondeforming qualities to make these steels very useful for dies. In general, high-carbon, high-chromium steels can be divided into two main classifications: those that are essentially oil hardening (D3) and those that are essentially air hardening (D2). These steels are widely used for blanking and cold forming of punches and dies.

2.7 *Air-hardening Cold Work Die Steels*

AISI A2 (1.0 C, 0.6 Mn, 0.25 Si, 5.0 Cr, 0.25 V; air hardening); AISI A6 (0.7 C, 2.0 Mn, 0.25 Si, 1.0 Cr, 1.25 Mo; air hardening).

Air-hardening die steels compete with both the manganese oil-hardening steels (O2, L6) and high-carbon, high-chromium steels (D2, D3) as cold work die materials. As their name implies, these steels possess deep air-hardening properties. They are extremely useful for intricate dies that must maintain their shape after hardening and tempering since the movement resulting from air hardening of these steels amounts to only about one-fourth of that of a manganese oil-hardening steel like O2. Their wear resistance is intermediate between the manganese types and the high-carbon, high-chromium types, but their toughness is greater than both. They machine with but slightly more difficulty than manganese oil-hardening steels when annealed.

The unique combination of properties found in air-hardening die steels makes them especially suitable where fair abrasion resistance must be combined with exceptional toughness; thus they are widely used for blanking and forming dies, rolls, punches, drawing dies, and thread rolling dies. In addition, they have application for certain types of shear blades. Because of the inherent advantages of air hardening, these steels are replacing oil-hardening tool steels to an ever-increasing extent.

2.8 Tungsten Die Steels for Hot Work

AISI H21 (0.35 C, 0.30 Mn, 3.5 Cr, 0.5 V, 9.0 W; red hard).

Tungsten steels were historically the first high-alloy die steels used for hot work tooling. They contain low percentages of carbon, and from 9% to 18% tungsten plus additions of other alloying elements. All of these steels contain 2.00–4.00% chromium, most of them have about 0.40% vanadium, and one of them contains 2.00% molybdenum. Being alpha-forming elements, they augment the effective tungsten content and lead to additional quantities of ferrite at elevated temperatures. In certain types addition of the gamma-forming elements nickel or cobalt counterattacks this tendency and permits higher quenched hardness with relatively lower carbon content.

The tungsten die steels possess greater hot hardness than any other class of hot work steels except the high-speed steels. However, their higher carbon content makes them less shock resistant than tungsten hot work die steels. Tungsten steels are not as resistant to thermal shock as chromium and chromium–molybdenum steels and therefore cannot be rapidly cooled with water during operation without danger of breakage. Thus, tungsten die steels are employed where maximum hot strength and resistance to shock, while important, is a secondary consideration. Typical applications for these steels include extrusion dies for brass, bronze, and steel, hot press dies, dummy blocks, hot swaging dies, and hot punches.

2.9 Chromium–Molybdenum Hot Work Die Steels

AISI H13 (0.35 C, 0.30 Mn, 1.0 Si, 5.0 Cr, 0.40 V, 1.50 Mo; red hard).

The steels in this class are the most widely used of all the hot work die steels. The 5% chromium steels were originally developed for the die casting of aluminum alloys. The requirements were that they should be air hardened from a relatively low temperature, that they show little movement in hardening, have a minimum tendency to scale on air cooling, resist heat checking caused by alternating heating and cooling, resist the erosive action of aluminum, and not be too high in alloy content to make the cost of the steel prohibitive.

The outstanding characteristic of these steels is their toughness, which distinguishes them from most other tool steels. Although their hot hardness is not as great as the higher-alloy hot work die steels, such as H21, their extraordinary shock resistance makes them preferable on most hot work tasks and especially when it is necessary to cool the dies in service with water or other flushing media.

Typical applications for these steels include die-casting dies, forging dies, punches, piercers and mandrels for hot work, hot extrusion tooling, shear blades for hot work, and all types of dies for hot work that involves shock. Certain of these steels are used for ultrahigh strength structural parts.

2.10 High-speed Steels

AISI M2 (0.85 C, 0.30 Mn, 0.30 Si, 4.0 Cr, 2.0 V, 6.0 W, 5.0 Mo; red hard).

High-speed steels are primarily used in the manufacture of cutting tools. The most important property of high-speed steels is the ability to retain a high hardness at elevated temperatures.

3. Conclusion

There are essentially three broad classes of tool steels: tool steels, die steels (both cold work and hot work), and the high-speed steels. This article discussed the effects of alloying elements on tool steel properties and structure in detail and then briefly discussed the types, properties, and selection of the various tool and die steels. For more information on tool and die steels see the bibliography.

Bibliography

Brick R M, Gordon R B, Phillips A 1965 *Structure and Properties of Alloys*, 3rd edn. McGraw-Hill, New York

Palmer F R, Luerssen G V, Pendleton J S 1978 *Tool Steel Simplified*, 4th edn. Chilton, Radnor, PA

Roberts G A, Cary R A 1980 *Tool Steels*, American Society for Metals, Metals Park, OH, 4th edn

Roberts G, Krauss G, Kennedy R 1998 *Tool Steels*, 5th edn. ASM International, Materials Park, OH

Thelning K E 1975 *Steel and its Heat Treatment*. Butterworths, London

P. M. Novotny
Carpenter Technology Corporation, Reading, Massachusetts, USA

This Page Intentionally Left Blank

W

Wood, Constituents of

Wood is a cellular lignocellulosic material consisting of cellulose embedded in a hemicellulose/lignin matrix. The composition of wood varies at all levels from species to species, among cell types, and within the cell wall itself. The chemical composition of softwoods (conifers) differs from that of hardwoods (arboreal angiosperms) in the structure and content of lignin and hemicelluloses (Table 1). Wood may also contain significant quantities of extractives (resins and gums) in the form of low molecular weight extracellular compounds. The inorganic component of wood is typically less than 0.5%.

1. *Cellulose*

Cellulose is one of the most abundant organic compounds on earth. It is a linear homopolysaccharide made up of β-1,4-linked D-glucopyranoside residues in the $^4C_1$ chair conformation (Fig. 1, (**1**)). In both softwoods and hardwoods the cellulose content is approximately 42%. In the native state, adjacent parallel cellulose chains are aggregated together, by inter- and intramolecular hydrogen bonds, to form microfibrils approximately 4nm wide. These microfibrils contain both highly ordered (crystalline) and less ordered (amorphous) regions. The native crystalline structure of cellulose is classified as the cellulose *I* allomorph. The measured degree of polymerization of native cellulose is of the order of 10 000. The microfibrils are further aggregated with lignin and hemicelluloses to form macrofibrils. The high level of bonding in cellulose micro- and macrofibrils results in both high tensile strength and general insolubility in most solvents.

2. *Hemicelluloses*

Classification of plant polysaccharides has been based on their extraction methods. Hemicellulose is a term used for plant cell wall polysaccharides, other than cellulose and pectin, which are soluble in alkali solutions. Hemicellulose consists of various heteropolysaccharides, which are deposited in the cell wall at a level of 30 ± 5%, and are closely associated with cellulose and lignin. They can be readily isolated, after delignification, by sequential alkali extraction of increasing concentration. In softwoods the predominant hemicellulose (20 ± 5%) is *O*-acetyl galactoglucomannan (Fig. 1, (**2**)). The polysaccharide is composed of a linear backbone of β-1,4-linked D-mannopyranosyl and β-1,4-linked D-glucopyranosyl residues in a ratio of approximately 4:1, respectively. Some of the mannosyl residues are substituted with (i) α-D-galactopyranosyl units at position C-6 and (ii) *O*-acetyl groups, predominantly at position C-3 and to a lesser extent at position C-2. The level of galactose substitution can vary, relative to glucose, from 0.1 to 1.0. The ratio of *O*-acetyl to mannose residues is about 1:6.

An acidic arabino-4-*O*-methylgluconoxylan hemicellulose fraction is also present in softwoods at about 10%. This heteropolysaccharide is made up of a β-1,4-linked D-xylopyranosyl backbone substituted with α-L-arabinofuranosyl and 4-*O*-methyl-α-D-glucuronopyranosyl side groups at positions C-3 and C-2, respectively (Fig. 1, (**3**)). Generally, the xylose to arabinose to uronic acid ratio is 6:1:1. In certain softwoods, such as larch, arabinogalactan II is readily extracted by hot water to give yields of between 5% and 30%, and is a highly branched polysaccharide built up of a β-1,3-linked D-galactopyransoyl backbone to which at position C-6 oligosaccharide side-chains are attached.

In hardwoods, *O*-acetyl-glucuronoxylan is the predominant hemicellulose (25 ± 5%) which has the same

Table 1
Chemical composition of normal and reaction woods in typical softwood and hardwood species.

	Radiata pine (*Pinus radiata*)		Birch (*Betula papyrifera*)	
Component	Normal wood (%)	Compression wood (%)	Normal wood (%)	Tension wood (%)
Cellulose	40	36	42	50
O-Acetylgalactoglucomannan	20	12		
Glucomannan			3	2
1,4-Galactan		8		8
Arabino-4-*O*-methylglucoronxylan	11	11		
O-Acetyl-4-*O*-methylglucoronxylan			30	22
Lignin	27	32	24	17
Extractives	2	1	1	1

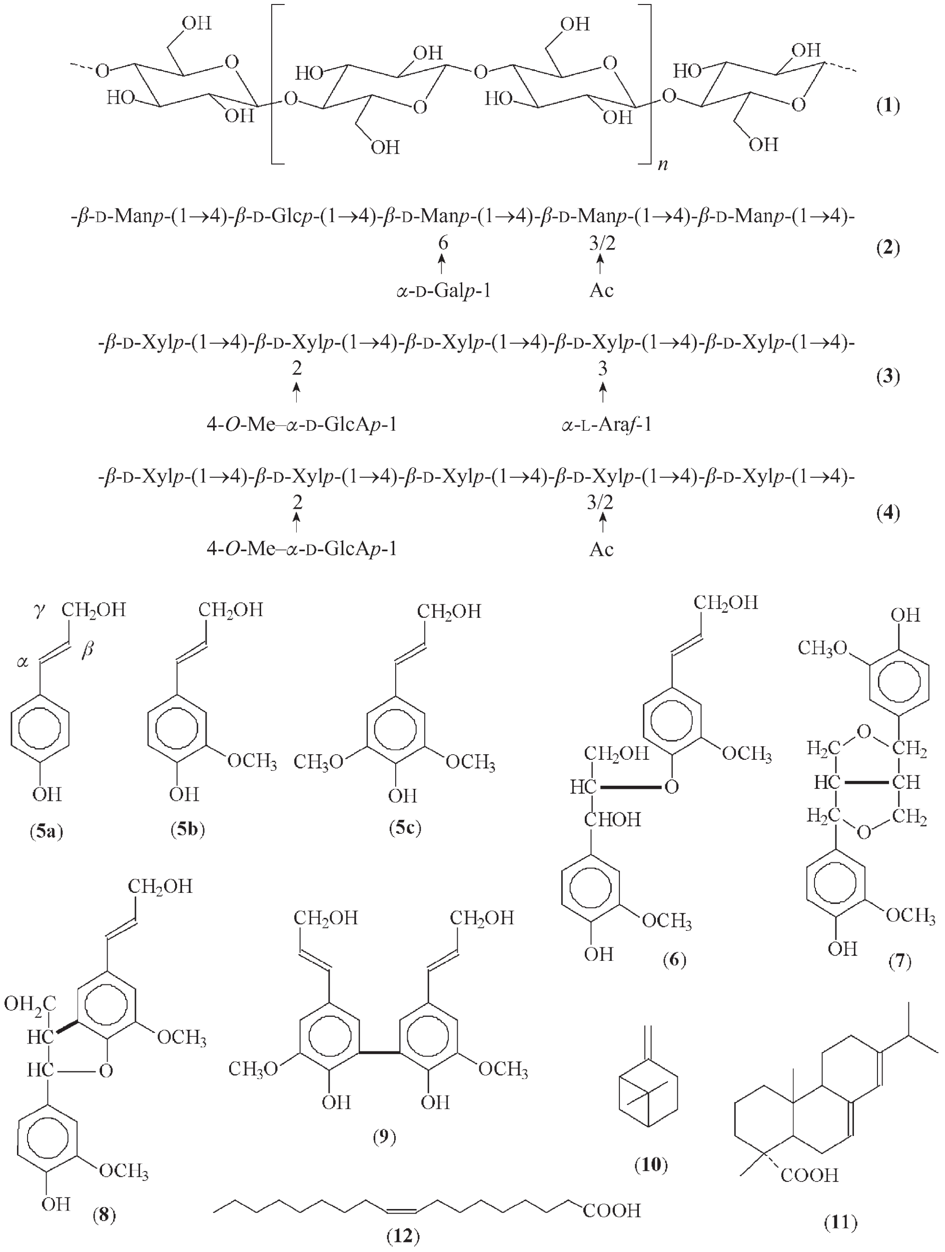

(1)
-β-D-Manp-(1→4)-β-D-Glcp-(1→4)-β-D-Manp-(1→4)-β-D-Manp-(1→4)-β-D-Manp-(1→4)-
6 ↑ α-D-Galp-1
3/2 ↑ Ac
(2)
-β-D-Xylp-(1→4)-β-D-Xylp-(1→4)-β-D-Xylp-(1→4)-β-D-Xylp-(1→4)-β-D-Xylp-(1→4)-
2 ↑ 4-O-Me-α-D-GlcAp-1
3 ↑ α-L-Araf-1
(3)
-β-D-Xylp-(1→4)-β-D-Xylp-(1→4)-β-D-Xylp-(1→4)-β-D-Xylp-(1→4)-β-D-Xylp-(1→4)-
2 ↑ 4-O-Me-α-D-GlcAp-1
3/2 ↑ Ac
(4)
(5a)
(5b)
(5c)
(6)
(7)
(8)
(9)
(10)
(11)
(12)

Figure 1
Common chemical structures found in wood.

main chain of a β-1,4-linked D-xylopyranosyl units, as in softwoods, and substituted with 4-*O*-methyl-α-D-glucuronopyranosyl and *O*-acetyl groups at positions C-2 and C-3/2, respectively (Fig. 1, (**4**)). Typically the xylose to acetyl to uronic acid ratio is about 10:7:1.

Furthermore, hardwoods contain a small amount of glucomannan (< 5%), which is composed of a linear chain containing β-1,4-linked D-mannopyranosyl and β-1,4-linked D-glucopyranosyl residues generally in a ratio 2:1.

3. Lignin

Lignin is a three-dimensional polymer built up of phenylpropane units that is laid down within the cell wall after tracheid elongation has ceased. The incorporation of lignin into the cellulose microfibril structures within the cell wall greatly enhances the mechanical strength properties of wood, over pure cellulose. Typically softwoods contain 30 ± 5% lignin while hardwoods generally have a lower lignin content (25 ± 5%). Lignin can be isolated by solvent extraction in reasonable yield from finely ground wood, after a cellulase pretreatment. Monomeric lignin precursors are *trans p*-coumaryl, coniferyl, and sinapyl alcohols (Fig. 1, (**5a**), (**5b**), and (**5c**), respectively) that undergo dehydrogenative polymerization by peroxidase and/or laccase activity to form macromolecular lignin by random coupling. The reactivity and levels of the lignin precursors govern the final constitution of lignin. The most predominant linkage between phenylpropane units in both softwood and hardwood lignin is the β-O-4 linkage (Fig. 1, (**6**); guaiacylglycerol-β-coniferyl ether; 50%), the β–β linkage (Fig. 1, (**7**); pinoresinol; 5%), the β-5 linkage (Fig. 1, (**8**); phenylcoumaran; 10%), and the α-O-4 linkage (10%). Softwood lignin also contains the 5–5 linkage (Fig. 1, (**9**); biphenyl; 10%). Softwoods contain lignin made up of guiacyl units, while hardwood lignin is built up from both guiacyl and syringyl units.

Production of wood pulp for the paper industry often involves the chemical removal of lignin, which is burned to drive the chemical recovery systems. A few lignin-derived chemicals are produced, such as vanillin and lignosulfonates.

4. *Extractives*

The term "extractives" means a mixture of compounds that are extracted from wood by a solvent or solvents. Lipophilic extractives are referred to as wood resin. Typically, extractives are of low molecular weight and are located outside the cell wall. Extractives, in part, contribute to the natural durability of wood by containing substances that are toxic to bacteria, fungi, and insects. In pulping, extractives can be collected as by-products such as tall oil and turpentine. However, extractives can have a negative effect on paper quality such as pitch deposits. Heartwood generally contains higher levels of extractives than sapwood. Extractives can be broken down by classes of compounds, such as monoterpenes (e.g. β-pinene; Fig. 1, (**10**)), resin acids (e.g. abietic acid; Fig. 1, (**11**)), fatty acids (e.g. oleic acid; Fig. 1, (**12**)), fats, sterol esters, phenolics, tannins, flavonoids, lignans, cyclitols, carbohydrates, and amino acids.

5. *Reaction Wood*

Reaction wood is produced as a response to stem lean, and has the function in the tree to generate growth stress to correct the stem lean, thus maintaining the most efficient position of the stem for photosynthesis. Reaction wood occurs in two different forms, compression wood in softwoods, and tension wood in hardwoods. Compression wood typically has more lignin, and less cellulose and galactoglucomannan, compared to normal wood, while tension wood has a higher cellulose content. Compression wood lignin has a higher content of *p*-coumaryl units. Both compression wood and tension wood contains a β-1,4-linked D-galactan not found in normal wood. The increased lignification of compression wood tracheids is thought to produce significant longitudinal swelling during wood formation, thus generating an expansion on the lower side of the stem, which attempts to correct the lean. In tension wood, the additional cellulose is thought to generate tension on the upper side of the stem due to shrinkage during wood formation. Unfortunately the growth stress which is beneficial to the tree has undesirable effects on timber properties resulting in abnormal shrinkage and dimensional instability.

6. *Topochemistry*

The wood cell wall has a complex structure with variations in chemical composition among different cell wall layers. Wood also contains different cell types, which in turn may have a different chemical composition. In softwoods, the tracheid accounts for 95% of the wood substance and thus has a dominant effect on the chemical composition. The tracheid cell wall consists of a highly lignified middle lamella incorporating the primary wall. The secondary wall is lignified to a much lesser extent than the middle lamella with higher lignin concentrations often present in the S3 layer lining the lumen. Both the inner S3 and outer S1 layers have a denser hemicellulose matrix than the S2 layer which is usually the thickest layer. An important effect of this complex variation in chemical composition across the cell wall is that chemical composition of a bulk wood sample can be altered significantly by changes in the proportion of the different cell wall layers. Chemical composition can also vary in proportion to wood density for the same reason. This often makes it difficult to distinguish true changes in chemical composition from changes resulting from anatomical or ultrastructural variations.

The middle lamella typically contains 20–25% of the total lignin content while the lignin concentration in this region varies from 50% to 80%. Interestingly, some of this variation has been shown to be genetically inherited among clones within a species. The secondary wall typically has a lignin concentration of 20–25%. In compression wood the distribution of lignin alters dramatically, with reduced lignification of the middle lamella and increased lignification of the secondary wall, especially in the outer S2 region at the corners of the tracheids known as the S2L region, where the lignin concentration may exceed that in the middle lamella. In tension wood the increased cellulose content is due to the presence of an additional cell wall layer consisting largely or entirely of cellulose known as the gelatinous layer because of its characteristic detached appearance under the light microscope.

In hardwoods, lignin varies in its chemical composition among cell types and cell wall regions. Guiacyl lignin predominates in cells involved with water conduction (vessels) while syringyl lignin predominates in fibers and parenchyma cells. The middle lamella region of vessels also contains guiacyl lignin while the middle lamella of fibers contains a mixed guiacyl/syringyl lignin.

7. *Conclusions*

The chemical composition of wood is highly variable and this variability has a significant effect on wood properties. Research is increasingly being aimed at the manipulation of chemical composition by genetic engineering and the use of molecular biology to study the process of wood formation. The driving force behind this research is the need to develop environmentally friendly processing options and to develop novel wood based products.

See also: Wood: Macroscopic Anatomy; Wood: Ultrastructure

Bibliography

Fengel D, Wegner G 1989 *Wood: Chemistry Ultrastructure, Reactions.* De Gruyter, Berlin

Higuchi T 1997 *Biochemistry and Molecular Biology of Wood.* Springer, Berlin

Timell T E 1986 *Compression Wood in Gymnosperms.* Springer, Berlin, Vol. 1

A. G. McDonald and L. A. Donaldson
Forest Research, Rotorua, New Zealand

Wood: Macroscopic Anatomy

Wood is a cellular material produced by a living tree to support the leafy crown, conduct water and dissolved nutrients from the roots upward to the crown, store reserve food (primarily carbohydrates) that is later used for tree growth, and to synthesize extractives during heartwood formation or in response to injury. The anatomy of wood reflects these biological functions. Support and conducting cells are usually "pipe-like"; they have rigid walls, generally are longer than they are wide, and normally are oriented with their long axes parallel to the long axis of the tree. Water-conducting cells are dead hollow cells, as are the support cells in most commercially important US woods. Storage cells (parenchyma) generally are very small, and are aggregated into horizontally oriented structures called rays. In addition, in many hardwoods and a few softwoods there may be longitudinally oriented storage cells. Reserve food materials are moved "in and out" of the wood via the rays. The anatomy of wood affects its material characteristics. For example, the interfaces between rays and the longitudinal elements seem to be "weak points" in wood. It is at these interfaces and within the rays that separations between cells (checks) originate during wood drying, and wood is easiest to split.

As one would expect for a substance of biological origin, wood is a variable material. The types of woody cells, their relative abundance, and arrangements differ among species. There also is within species a variation in cell structure related to growing conditions, and within trees variation related to tree age. Major challenges for the forest products industry have been, and continue to be, the ability to grow wood that is predictable in its behavior and to match wood characteristics to product requirements.

1. *Wood Types: Hardwoods and Softwoods*

In the forest products industry, the terms softwood and hardwood are applied to two major groups of trees that correspond to two major botanical groupings. Softwood trees have seeds that are not covered, but are produced in cones (conifers such as pine, redwood, cedar, cypress). Hardwood trees produce covered seeds within flowers (dicotyledonous angiosperms such as oak, ash, maple, teak, mahogany). The use of the terms softwood and hardwood is somewhat unfortunate as it implies absolute differences between the two groups in the hardness of their woods. However, the wood of some hardwoods is softer than that of some softwoods (and vice versa). One of the world's lightest woods, *Aeschynomene hispida*($30\,\mathrm{kg\,m^{-3}}$), comes from a dicotyledonous tree, and, in that sense, is a hardwood. For commercially important woods native to the USA the density of softwoods is $290–600\,\mathrm{kg\,m^{-3}}$, average $410\,\mathrm{kg\,m^{-3}}$; of hardwoods is $320–810\,\mathrm{kg\,m^{-3}}$, average $500\,\mathrm{kg\,m^{-3}}$. Density is more variable within tropical woods than it is within temperate-zone woods.

The wood structure of hardwoods and softwoods is distinct, with softwoods having the simpler wood anatomy. In softwoods, a single cell type, the longitudinal tracheid, functions for both support and conduction, and constitutes 90% of the wood volume (Fig. 1). Longitudinal tracheids are usually 100 times as long as they are wide, with lengths usually of 2–4 mm, and up to 8 mm in large softwood trees, such as redwood (*Sequoia*). Rays in softwoods account for no more than 10% of the wood volume, and are very narrow.

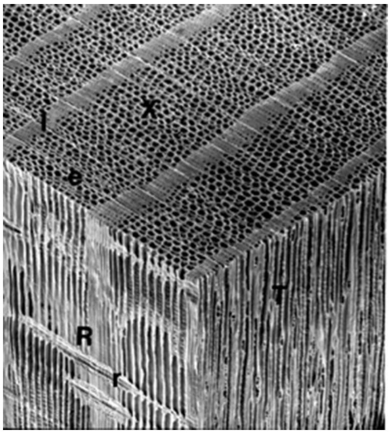


Figure 1
Scanning electron micrograph of softwood block: X, cross-section; e, earlywood with longitudinal tracheids with relatively wide openings; l, latewood with longitudinal tracheids with narrow openings; R, radial section; r, ray extending across growth ring boundaries; T, tangential section (courtesy of N. C. Brown Center for Ultrastructure Studies, SUNY College of Environmental Science and Forestry).

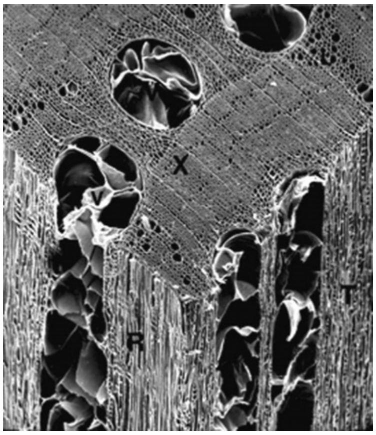


Figure 2
Scanning electron micrograph of ring porous hardwood block: X, cross-section; T, tangential section; R, radial section; V, vessel. Note the wide vessels (V) that are filled with tyloses. In the conducting sapwood the vessels would be hollow. In this wood, the tyloses fill the vessels and reduce permeability (courtesy of N. C. Brown Center for Ultrastructure Studies, SUNY College of Environmental Science and Forestry).

Hardwoods have vessels for water conduction, and fibers for support (Fig. 2). Vessels are composite structures composed of open-ended vessel elements that have relatively large diameters, thin walls, and large central openings (lumens). Individual vessel elements are usually 25–300 μm wide and 0.2–1 mm long. The composite vessels range from a few centimeters to more than 10m long. Fibers are narrow tubes (usually 10–30 μm wide and 1–2 mm long) with varying cell wall thickness. Support and conducting cells usually are dead at functional maturity, although in some species fibers are living in the sapwood and are involved in food storage. Ray size varies between hardwood species. In commercially important hardwoods, rays usually account for 10–30% of the wood volume, and vary from 1 cell wide and not visible to the eye (e.g., willow (*Salix* spp.) and cottonwood (*Populus* spp.)), to more than 20 cells wide and visible on all surfaces of the wood (as in oak (*Quercus* spp.)).

Proportions and arrangements of the different cell types, as well as cell wall thickness, differ from species to species. Woods that have a high proportion of thick-walled fibers are hard and heavy. Hickory (*Carya* spp.) is such a wood, and is the preferred material for tool handles in the USA. Woods with a high proportion of vessels and/or thin-walled fibers are soft and light. Balsawood (*Ochroma* spp.) is such a wood, and is used for model airplanes, and as an insulator because of the high proportion of open lumens.

2. *Cell Orientation*

The orientation of the woody cells within the trunk is of major importance in explaining many of the properties of wood. Cells that function in support and conduction, and, at times, some parenchyma cells, have a longitudinal orientation, i.e., the long axes of the cells are parallel to the long axis of the tree. Grain refers to the orientation of the support and conduction cells. In straight-grained wood, the support and conducting cells are oriented parallel to the outside of the tree; in spiral-grained wood, the support and conducting cells literally spiral (right-handed, left-handed, or alternating between the two) around the tree trunk. Some species (e.g., sweetgum (*Liquidambar styraciflua*)) are more apt to produce spiral grain than others. Straight-grained wood is preferred for most applications, as it is easier to dry without warping and twisting. However, in some mahogany trees, alternating spiral grain imparts a pleasing appearance to the wood, and such ribbon-striped mahogany is valued as a decorative veneer and for inlays.

Because wood is composed of cells oriented longitudinally and horizontally, it will have a different appearance depending on how it is cut, as different aspects of the cells are exposed (Fig. 3). A cross-section is produced when the cut is at right angles to the long axis of the tree. When a tree is cut down, the end of the stump reveals a cross-section. In this section, growth rings are evident as a series of concentric circles, and rays appear as fine lines crossing the

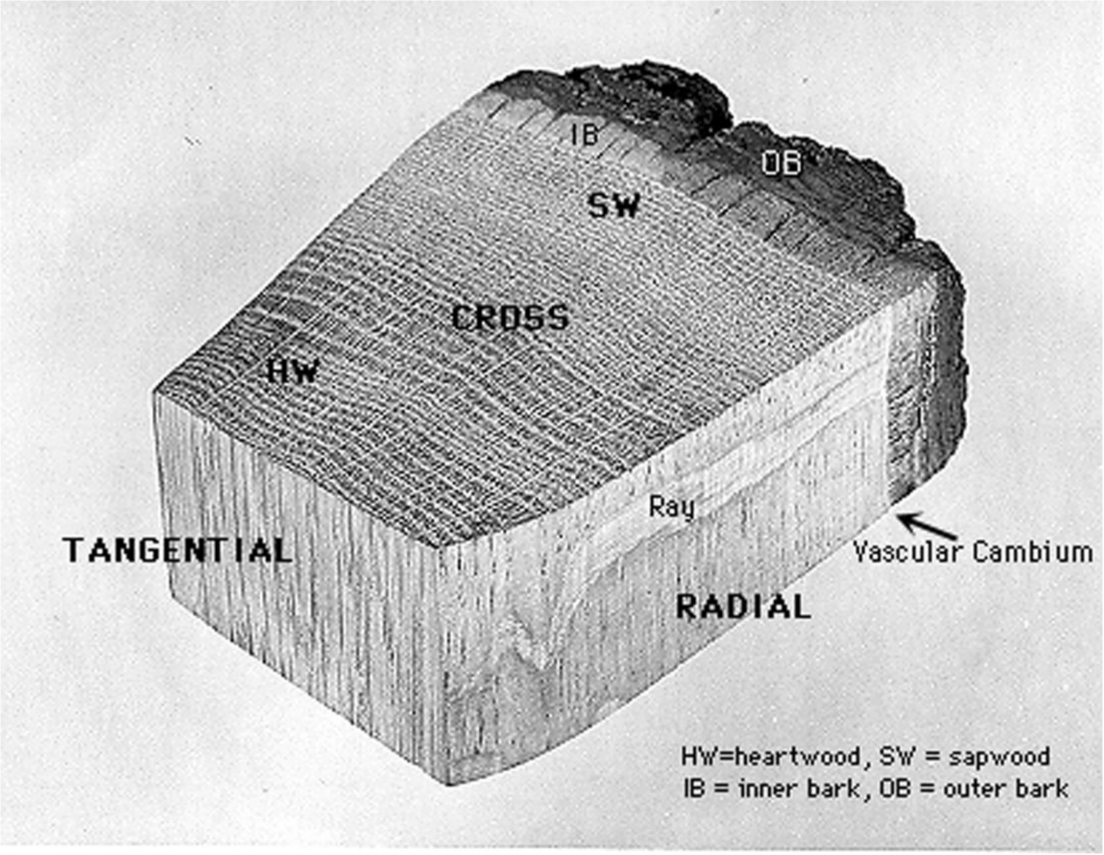


Figure 3
Block of red oak (*Quercus* sp.) showing bark (OB, outer bark composed of alternating layers of cork and dead food-conducting cells; IB, inner bark with functional food-conducting cells) and wood (SW, sapwood, in oak usually 10–12 outer rings; HW, heartwood, the region where the storage cells are dead). Cross, radial, and tangential surfaces are labeled as such, and the broad rays of oak are visible in all three surfaces.

growth rings at right angles. Longitudinal sections are produced by cutting parallel to the long axis of the tree. There are two types of longitudinal sections: radial and tangential. When the longitudinal cut is made by cutting parallel to the rays and perpendicular to the growth rings, a radial section is produced. In radial sections rays are viewed from the side, look like stripes, and have a "brick-wall" appearance with the individual cells in a ray being equivalent to individual bricks in the wall. Hard maples and oaks have a distinctive ray fleck on their radial surfaces. A tangential section is produced by cutting parallel to growth ring boundaries, or the surface of the tree, and cutting across the rays at right angles. The ends of rays are visible in tangential sections; ray height and width can be determined from tangential sections. Lumber cut along the radial section is termed quarter-sawn and that cut in the tangential plane is flat-sawn.

Wood behaves differently along the different axes. Water is lost most readily from the cross-section, and penetration of fluids (e.g., preservatives, pulping liquor) is easiest into the cross section. It is easiest to split wood longitudinally, particularly along rays, as it is easier to split cells apart, rather than cut across their cell walls. Most of the mechanical properties of wood are greatest along the grain, while shrinkage and swelling with changes in moisture content are less along the grain.

3. Growth Rings

New wood and new inner bark are inserted between existing wood and inner bark by cell divisions within the vascular cambium, a layer that is considered to be but one cell wide. Stripping the bark from around a tree can seriously injure (or kill) it, as usually the vascular cambium will also be stripped off, and the ability of the tree to form new conducting tissues is seriously compromised.

When trees grow in seasonal climates, they almost always produce trunk wood with growth rings. Cell production varies with the season, and there often are differences in the texture of the wood formed during different seasons. The wood formed at the beginning of the growing season is the earlywood (in north temperate regions often referred to as springwood); the wood formed later in the season is the latewood (in north temperate regions often referred to as summerwood). The latewood generally is denser and darker than the earlywood. Growth rings are visible because of the difference in texture between the latewood (usually comprised of relatively small and thicker-walled cells) and the earlywood of the subsequent year (with relatively large and thin-walled cells). In the north temperate region, a growth ring usually represents the amount of wood accumulated during one year and so is often called an annual ring. In the tropics, growth rings can be distinct, indistinct, or absent. If growth rings are present, they do not necessarily represent the growth of a single year, since growth often reflects wet and dry seasons. Determining the age and growth rate of tropical trees is more difficult than for temperate woods because growth rings do not always correspond to annual rings.

The appearance and properties of wood are affected by the ratio of earlywood to latewood, and how well defined the differences are in the transition between these zones. Diffuse porous hardwoods do not have a pronounced difference between earlywood and latewood, and the size and spacing of the vessels is nearly the same throughout a growth ring (Fig. 4). Almost all tropical hardwoods are diffuse porous. Ring porous hardwoods have earlywood with large diameter vessels, and a distinct latewood zone with small diameter vessels (Fig. 4). Approximately 25% of the hardwoods from the north temperate zone are ring porous. Some woods are intermediate between diffuse porous and ring porous and have a gradual change in vessel diameter throughout the growth ring (e.g., walnut (*Juglans* sp.)), and are called semi-ring porous.

Growth rate affects the density and openness of ring porous woods (e.g., oak (*Quercus* spp.), ash (*Fraxinus* spp.), hickory) more than it affects diffuse porous woods (e.g., birch (*Betula* spp.), maple (*Acer* spp.)). In ring porous woods, within a growth ring there is an earlywood zone that is of near-constant width each year, while the width of the latewood zone varies. Slow-grown ring porous woods with narrow growth rings will have proportionally more earlywood with a high vessel volume than latewood per unit area, and so will have a lower density than rapid-grown ring porous woods that will have proportionally more latewood with a high fiber volume and, consequently, a higher density.

Softwoods also vary in the proportions of earlywood and latewood, and the degree of difference between the two regions. Some species have an abrupt transition from the earlywood to the latewood, and a distinct band of latewood is formed (e.g., hard pines such as loblolly pine (*Pinus taeda*)). Others (e.g., soft pines such as eastern white pine (*P. strobus*)) have a gradual transition from earlywood to latewood, and not much latewood at all. The relationship between ring width and density in softwoods varies; in some species growth rate does not affect density, while in others an increase in growth rate can result in either a decrease or an increase in density.

Cell size and variation therein affect wood texture. Hardwoods with narrow vessels and fibers, and softwoods with narrow tracheids are fine-textured; hardwoods with large diameter vessels, and softwoods with wide tracheids are coarse-textured. The hardwoods boxwood (*Buxus* sp.) and dogwood (*Cornus florida* L.) and the softwood yew (*Taxus* spp.) are noted for their even, fine texture and even wear during use. Ring porous hardwoods and softwoods with a distinct latewood band have an uneven texture, because of the pronounced difference in earlywood and latewood cell sizes. Some textures are correlated with particular regions of the world; hardwood trees with diffuse porous wood and large diameter vessels (more than 200 μm across) only occur in the tropics.

4. *Sapwood and Heartwood*

The outer wood with living parenchyma cells is sapwood. The number of years that the parenchyma cells live and therefore the width of the sapwood zone varies from species to species, but parenchyma cells eventually die, and the wood with only dead cells is termed heartwood. Often, but by no means always, the heartwood is darker colored than the sapwood. Before dying, parenchyma cells often manufacture extractives. These extra chemicals may color the wood, coat cell walls and passages between cells and thus lower permeability, change the affinity of the wood for water, alter its dimensional stability, and affect its durability (resistance to decay). For instance, heartwoods of redwood, cedar, and teak are noted for their durability and are used for outdoor furniture and shingles, but the sapwoods of redwood and cedars with very low extractive content are not resistant to decay.

4.1 Tyloses

During normal heartwood formation or after an injury, parenchyma cells adjacent to a vessel may ‘‘grow into’’ that vessel. This occurs usually in hardwood species where the interconnections between the vessels and the storage cells (ray parenchyma) are large. These infillings of the vessels by parenchyma

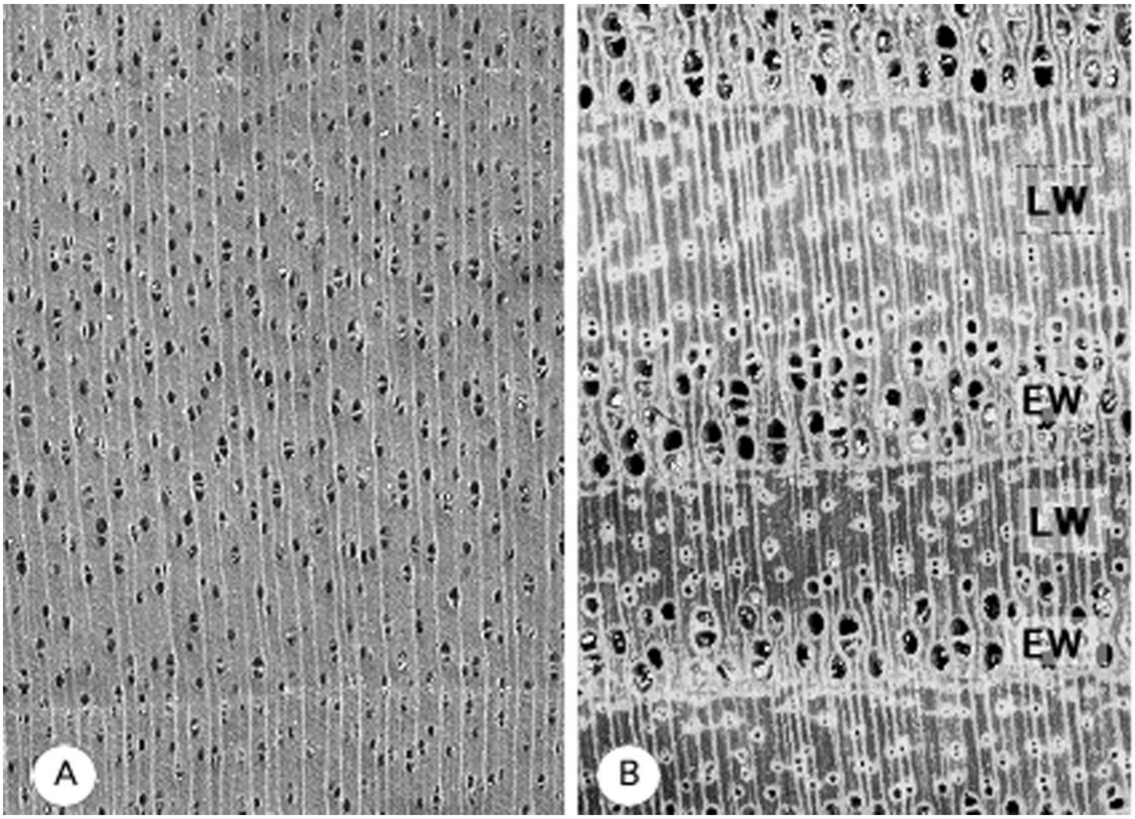


Figure 4
(A) Birch (*Betula* sp.), an example of a diffuse porous hardwood; vessel size and distribution is similar throughout the growth ring. (B) Ash (*Fraxinus* sp.), an example of a ring porous hardwood. The earlywood zone (EW) has large diameter vessels and the latewood (LW) has narrow vessels. Earlywood width does not vary much from year to year, but latewood width varies depending on growing conditions.

cells are called tyloses, and they lower the permeability of the wood. Woods with tyloses are slower drying and more difficult to treat with preservatives than woods without tyloses. White oak is noted for having many tyloses and a very low permeability.

5. Reaction Wood

It is termed abnormal wood, but the ability to form reaction wood to maintain the vertical alignment of a tree is important for the long-term survival of tree species. Reaction wood is of widespread occurrence. In hardwoods, reaction wood is formed on the upper side of a bend or lean and so is termed tension wood. If the growth rings are eccentric it is likely that tension wood is present. Tension wood often is difficult to recognize without microscopic examination, and can be found in straight stems. Tension wood fibers have a different wall structure than normal wood fibers. This variation is correlated with tension wood twisting and warping during drying, and being difficult to plane to a smooth surface. In extreme cases, logs with large amounts of tension wood will split immediately upon being sawn.

In softwoods, reaction wood is compression wood, and formed on the underside of a bend or lean. Compression wood also warps during drying, and is brash, breaking suddenly under load. It is also less permeable than normal wood, and has shorter cells. Unfortunately, pines are prone to form compression wood. Minimizing and eliminating compression wood improves the quality of softwoods for solid wood products as well as for pulpwood.

See also: Wood, Constituents of; Wood: Ultrastructure

Bibliography

Haygreen J G, Bowyer J L 1997 *Forest Products and Wood Science. An Introduction*, 3rd edn. Iowa State University Press, Ames, IA

Hoadley R B 1980 *Understanding Wood. A Craftsman's Guide to Wood Technology*. Taunton Press, Newtown, CT

Jane F W (revised by Wilson K, White D J B) 1970 *The Structure of Wood*. Adam & Charles Black, London

Panshin A J, de Zeeuw C 1980 *Textbook of Wood Technology: Structure, Identification, Properties, and Uses of the Commercial Woods of the United States and Canada*, 4th edn. McGraw-Hill, New York

Wilson K, White D J B 1986 *The Anatomy of Wood: Its Diversity and Variability*. Stobart & Son, London

E. Wheeler
North Carolina State University, Raleigh, North Carolina, USA

Wood: Ultrastructure

Other articles describe wood composition and cellular structure on a macroscopic scale, but in many instances the successful processing of wood requires a more in-depth understanding of these characteristics. A brief overview of wood chemical composition is given to provide a better understanding of cellular ultrastructure.

1. *Chemical Composition of Wood*

Wood is composed of three primary constituents: cellulose, a linear, crystalline carbohydrate polymer with a high degree of polymerization (~10000–14000); hemicelluloses, a term used to collectively refer to the numerous pentosan and hexosan carbohydrate polymers having a much lower MW and a greater solvent reactivity than cellulose; and lignin, a highly cross-linked polyphenolic material.

Two so-called secondary constituents are also present in, but not an integral part of, the cell wall structure: ash (salts of minerals such as calcium, potassium, and magnesium) and extractives (any compound that is removable with a neutral solvent). Ash content is usually $<0.5\%$ for all North American species, but may be as much as 5% or more for tropical species. Silica is sometimes present in appreciable quantities. Extractives frequently add color, odor, or other characteristics such as decay resistance to wood e.g., chemicals such as oleoresins (pine "pitch") and tannins. Different species contain different types of extractives, and wood from different parts of a tree will likely contain varying amounts of extractives e.g., heartwood usually has a greater extractive content than sapwood.

Table 1 lists the amounts of cellulose, hemicelluloses, and lignin in wood compiled from the literature on North American gymnosperms (conifers, "softwoods") and angiosperms (broad-leaved trees, "hardwoods") (Browning 1963, Timell 1967, Koch 1972, 1985, Page *et al.* 1975, Panshin and de Zeeuw 1980, Weil *et al.* 1998). Q3–Q1 represents the inner quartile spread. Tropical angiosperms may have significantly more lignin than temperate-zone angiosperms (Timell 1967). A study of 20 tropical woods (mostly from Nigeria) found that lignin content ranged from 22% for iroko (*Chlorophora excelsa*) to 50% for mpingo (*Dalbergia melanoxylan*) (Sosanwo *et al.* 1995). Cellulose contents were ~35%.

There are statistically significant differences between the hemicelluloses and lignin determinations for temperate-zone angiosperms and gymnosperms, the latter having less hemicelluloses and more lignin. Variability could be from differences in analytical procedures, but it is known that trees having different genetic heritage contain differing proportions of the primary constituents, and there is also variation within trees. In gymnosperms, for example, the juvenile wood surrounding the pith of the tree has more lignin and less cellulose than wood just beneath the bark formed from a more mature cambium. There is a similar trend with tree height; topwood (with a higher juvenile wood percentage) generally has less cellulose and more lignin than bottomwood. Hemicelluloses seem to be more concentrated in the juvenile regions as well, and usually decrease by a few percent in mature wood regions (Panshin and de Zeeuw 1980).

2. *Cell Wall Structure*

There are four essential relationships in the structure of wood cell walls: (i) aggregation of cellulose, lignin, and hemicelluloses molecules to form larger structures; (ii) arrangement of these structures to form a

Table 1
Primary constituents of wood (%) (moisture-, extractive- and ash-free basis).

	Cellulose	Hemicelluloses	Lignin
Gymnosperms			
Medians of 28 analyzes (Q3–Q1)	43.0 (7.7)	25.0 (11.7)	28.6 (2.9)
Minima (maxima)	38.1 (57.4)	13.3 (34.1)	23.5 (33.9)
Angiosperms			
Medians of 62 analyses (Q3–Q1)	45.8 (4.7)	31.1 (6.5)	26.4 (3.6)
Minima (maxima)	38.5 (55.2)	18.7 (39.7)	17.3 (34.7)

multilayered cell wall; (iii) variation of concentrations of cellulose, hemicelluloses, and lignin within the cell wall and the intercellular domain; and (iv) specialization of cell walls, such as interconnecting pits. Reaction wood has an abnormal cell wall structure that is described later.

2.1 Aggregation of the Primary Constituents into Larger Structures

As stated previously, cellulose is the principal constituent in cell walls. Recent work suggests that cellulose molecules are formed at macromolecular terminal complexes (TCs) in the cell plasma membrane. There is a multistage self-assembly sequence. Parallel glucan chains form planar sheets that quickly hydrogen bond to other sheets to produce mini-crystals, and associate with others to form highly-ordered cellulose microfibrils. Microfibrils are approximately square, and the atomic force microscope (AFM) has shown them to have repeat patterns along their length corresponding to the lengths of a glucose molecule and a cellobiose unit (two glucose molecules) (0.52 and 1.04 nm, respectively) (Hanley *et al.* 1992, Baker *et al.* 1998).

Measurements of bacterial cellulose and beaten kraft-pulped spruce fibers have shown that the lateral microfibril dimensions can be as small as 1.0–1.5 nm (Haigler and Weimer 1991, Okuda *et al.* 1994, Hanley 1995, Tsekos *et al.* 1996). Microfibrils have not been observed to be this small in undelignified structures. Other workers have demonstrated that the space between cellulose microfibrils is ~3–4 nm (Terashima *et al.* 1993) and that hemicelluloses at least partially cover the cellulose in native wood (Tenkanen *et al.* 1999). Hanley (1995) found that the widths of microfibrils from microtomed sections of undelignified black spruce were ~7.5 nm in diameter, probably an indication of the combined cellulose-hemicelluloses dimension. Lignin is deposited after the cell wall formation is initiated and does not affect cellulose dimensions or structure.

The wood cell wall structure is frequently thought of as a two-phase system in which cellulose is a discrete component and the lignin and hemicelluloses are blended together to form an encrusting matrix around the cellulose. Figure 1 illustrates this two-phase model, which has recently been questioned by some researchers. Studies using the cellulose-producing bacteria *Acetobacter xylinum* as a model system have shown that cellulose and hemicelluloses coaggregate to form complex celluloses. It is possible that hemicelluloses are blended with the cellulose in most higher plant cell walls (particularly in wood), and that microfibrils that are generally observed in isolated celluloses are consequences of the extraction of the hemicelluloses. Work by Salmén and Olsson (1998) used dynamic mechanical analysis to study spruce specimens at varying relative humidities, which indicated that hemicelluloses may be associated with both lignin and cellulose. According to their findings, xylan is associated with lignin and glucomannan with cellulose in spruce (*Picea abies*).

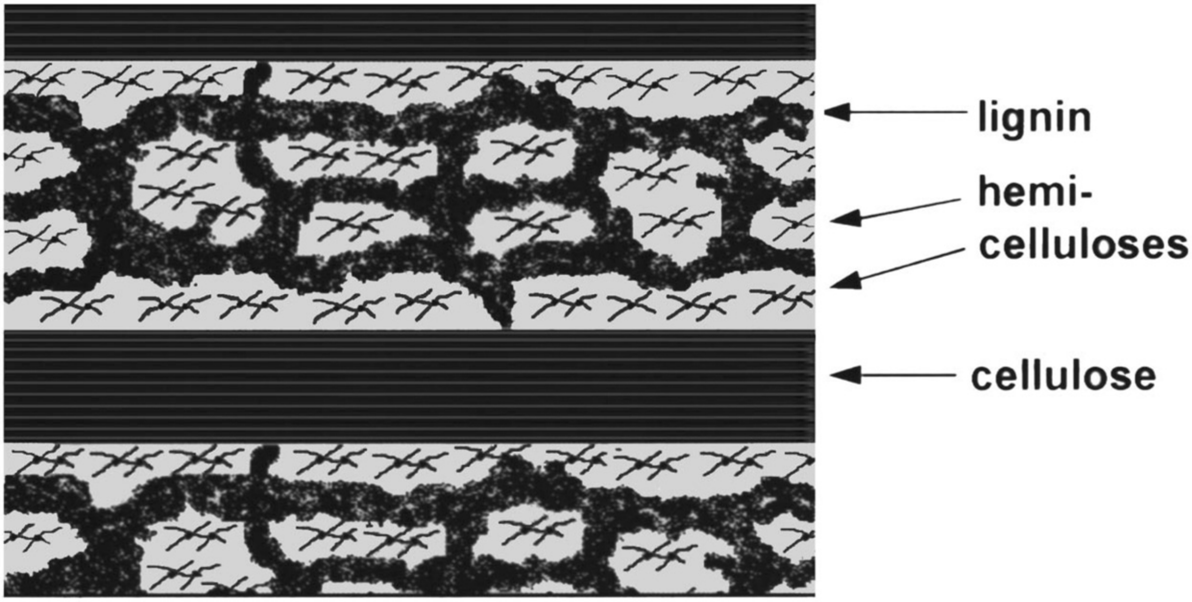


Figure 1
Two phase model for organization of the cell wall components (after Salmén and Olsson 1998).

Molecular mechanics and dynamics studies indicate that the hemicelluloses are hydrogen bonded and oriented parallel to cellulose, and that lignin may have a helical structure (Faulon *et al.* 1994). However, the interactions are complex and many inferences remain to be verified.

X-ray diffraction studies have demonstrated departures from an infinitely extended linear lattice in native cellulose microfibrils; consequently it has often been stated that microfibrils contain both crystalline regions and less well ordered (amorphous) regions. Raman spectroscopy and solid-state ^{13}C-NMR studies have shown that this model is inadequate to describe cellulose structure. Through use of the alga *Valonia*, and other native celluloses as experimental substrates it has been shown that two different forms of cellulose (I_α and I_β) are present in the same microfibril. There is speculation that the I_α and I_β forms of cellulose may correspond to triclinic and monoclinic unit cells. In wood, the concomitant presence of these two forms probably affects x-ray crystallinity measurements to a lesser degree than cellulose chain ends or the intimate associations with hemicelluloses. There is much yet to learn about nanoscale level of organization of constituents in wood.

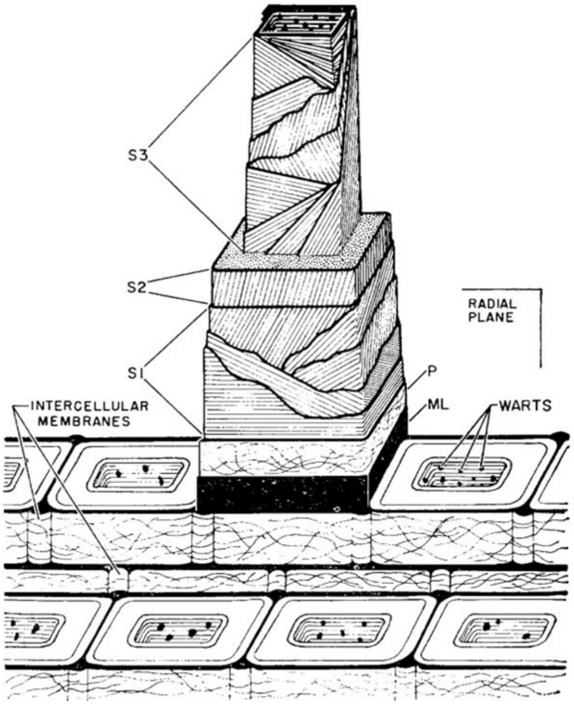


Figure 2
Cell wall structure of a longleaf pine (*P. palustris*) latewood tracheid (after Dunning 1968).

2.2 Multilayer Structure of Cell Walls

Wood cells are hollow layered structures with a primary cell wall on the outside and a secondary cell wall and a lumen contained within. Each layer contains oriented cellulose microfibrils together with hemicelluloses and lignin, and the microfibril orientation affects the ultimate strength of the tissue. The primary cell wall (P) is very thin and contains randomly oriented microfibrils surrounded by a lignin-rich material containing lesser amounts of cellulose, hemicelluloses, pectins, and structural glycoproteins. The secondary cell has layers visually distinguishable by differences in microfibrillar orientation; progressing from the primary cell wall inward towards the lumen, the layers are sequentially designated as S_1, S_2, and S_3. Figure 2 is a generalized diagram of a woody cell wall based on latewood tracheids of longleaf pine (*Pinus palustris*) (Dunning 1968).

The S_1 layer is slightly thicker than the primary cell wall, and the microfibrils are arranged in several thin lamellae with the outermost lamellae oriented at angles approaching 90° to the fiber axis. Two inner lamellae of the S_1 layer are arranged in alternating right- and left-hand helices at angles of $\sim 65°$ from the fiber axis, with several overlapping lamellae at progressively steeper orientations that create a gradual transition to the S_2 layer. The S_2 is the thickest layer, typically comprising around 75% of the cell wall, with the microfibrils arranged in a right-hand helix of 5–25° to the fiber axis. The S_3 layer is slightly thinner than the S_1, and similarly has several criss-crossed lamellae. The microfibril angle within the S_3 layer varies from $\sim$60–85°, with the outermost S_3 lamellae having similar angles as the S_2, acting as transitional elements between the S_2 and S_3. Finally, most gymnosperms (and some angiosperms) contain a thin "warty layer" on the innermost lamella of the S_3, the surface of the cell lumens. The warty layer is $\leqslant 0.1\ \mu$m thick and characterized by wart-like protuberances, which may be remnants of protoplasmic contents, and have no effect on fiber strength, although they may affect permeability.

Surrounding the primary cell wall is a thin lignin-rich layer termed the middle lamella (or inner cellular layer). In early microscopic studies these layers could not be clearly distinguished and the term compound middle lamella arose to refer to the combination of the (true) middle lamella and the primary cell wall. Table 2 lists the approximate thickness of the various cell wall layers of gymnosperm tracheids. Interested readers are referred to Dunning (1968) for more detailed information and micrographs of cell wall structures.

2.3 Distribution of Chemical Constituents

The distribution of the primary chemical constituents within wood cell walls has a significant effect on wood processing operations, and investigations have been ongoing since the seminal work by Bailey (1936). Studies have relied upon a number of destructive and nondestructive techniques e.g., microdissection, acid digestion, polarizing light microscopy, UV-microspectroscopy, enzyme-gold complex studies, Raman spectroscopy, bromine-tagged lignin measurements using x-ray spectroscopy, etc. All of these techniques have some limitations, and some contradictory results have been published that are difficult to resolve. General agreement exists on the qualitative distribution of cellulose, hemicelluloses, and lignin in the middle lamella and the secondary cell wall, but there is considerable disagreement on quantitative values. Table 3 and Fig. 3 show one possible distribution of cellulose, hemicelluloses, and lignin within the various cell wall layers of gymnosperms based on reports from Meier (1961), Fergus *et al.* (1969), Wood and Goring (1971), Saka *et al.* (1978), Hardell and Westermark (1981), Donaldson (1987), and Westermark *et al.* (1988).

When the nominal cell wall layer thicknesses listed in Table 2 are used to calculate the overall chemical composition, the results compare favorably to those presented in Table 1 (i.e., about 42% cellulose compared to 43% in Table 1, 32% hemicelluloses vs. 25%, and 25% lignin vs. 29%). In spite of the greater concentration of lignin in the compound middle lamella, it must be apparent that most of the lignin is actually present within the secondary cell wall due to its greater thickness.

Table 2
Cell wall layer thicknesses of gymnosperm tracheids (μm)

Middle lamella	$\geqslant 0.1$
Primary cell wall	0.03
S_1	0.2
S_2	$\geqslant 5$
S_3	0.07–0.08
Warty layer	$\leqslant 0.1$

Source: Kollmann and Côté (1968)

2.4 *Cell Walls and Distinguishing Features*

Up to this point the woody cell wall has been described in terms of chemical composition and structural organization of the principal constituents in the cell wall. However, wood cells are not balloon-like bodies with smooth walls and closed ends, but have many features that characterize the cell walls. Foremost are the paired openings in adjacent cell walls that permit the passage of fluids. In both gymnosperms and angiosperms some cell walls are penetrated by openings called pits. Angiosperms also have openings (perforation plates) at the junctures between ends of individual vessel elements. Other features are less uniformly present and are of diagnostic significance to identify wood species.

(a) Gymnosperm pits

In gymnosperms pits are found in three locations: intersections of longitudinal parenchyma and ray parenchyma cells, radial walls of adjacent longitudinal tracheids, and intersections of longitudinal

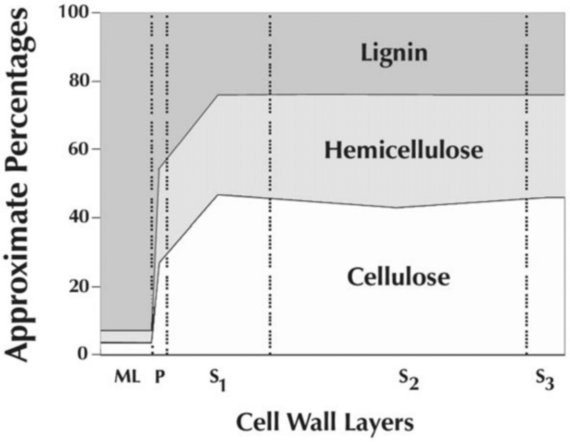


Figure 3
Possible cellulose, lignin and hemicellulose distribution in a gymnosperm cell wall.

Table 3
Gymnosperm cell wall layer constituent (%).

Cell wall layer	Cellulose	Hemicelluloses	Lignin
Middle lamella	3.5	3.5	93
Primary cell wall	27	27	46
Compound middle lamella	10	10	80
S_1	47	29	24
S_2	43	33	24
S_3	46	30	24

tracheids with ray parenchyma or ray tracheids. Simple pits between adjacent parenchyma cells are thin, membrane-like structures comprising the middle lamella and the primary cell walls. A bordered pit pair has overarching structures on the lumen side of each cell wall surrounding apertures between two adjacent longitudinal tracheids (Fig. 4). Each bordered pit pair encloses a web-like microfibril structure called a *margo* that supports (except in western redcedar (*Thuja plicata*)) a central and much less permeable *torus*, the assembly acting as a valve to open or close the openings between tracheids (Fig. 5). When the torus is moved enough to block the pit aperture, as by surface tension during wood drying, the pit is said to be aspirated (Fig. 6). Longitudinal tracheids and ray tracheids are also interconnected via bordered pits, but the pits are much smaller than those between longitudinal tracheids.

Pits occurring at the crossfields between ray parenchyma and longitudinal tracheids are termed half-bordered pits because they have the characteristics of bordered pits on the longitudinal tracheid and simple pits on the ray parenchyma cell. These pits can be classified into five categories by differences in the shapes of apertures and overhanging borders: fenestriform (also called windowlike), pinoid, piceoid,

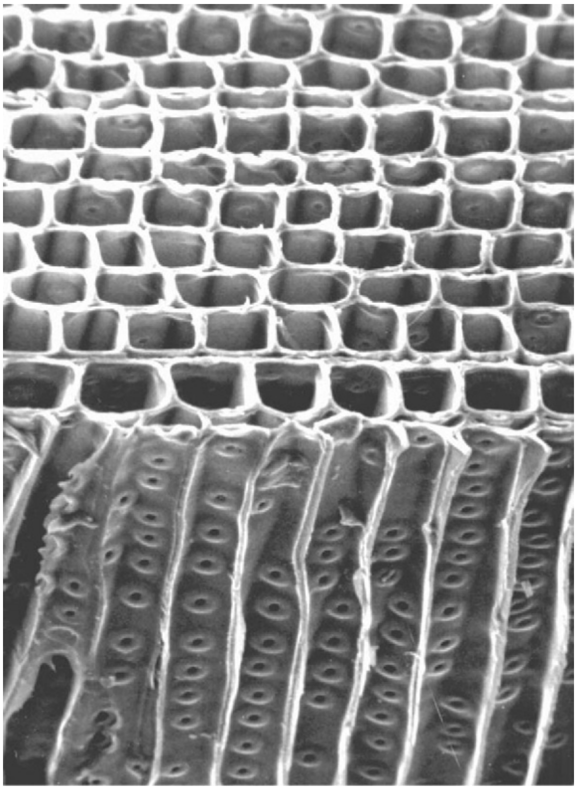


Figure 4
Longitudinal tracheids showing extensive bordered pitting on the radial cell walls of Jack pine (*P. banksiana*) (courtesy of the Institute of Paper Science and Technology).

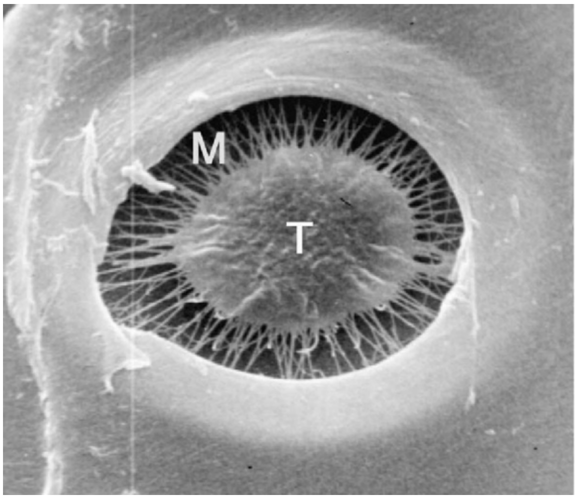


Figure 5
Margo (M) and torus (T) from an earlywood tracheid of balsam fir (*Abies balsamea*); the lumen-side pit border has been removed in this micrograph (courtesy of the Institute of Paper Science and Technology).

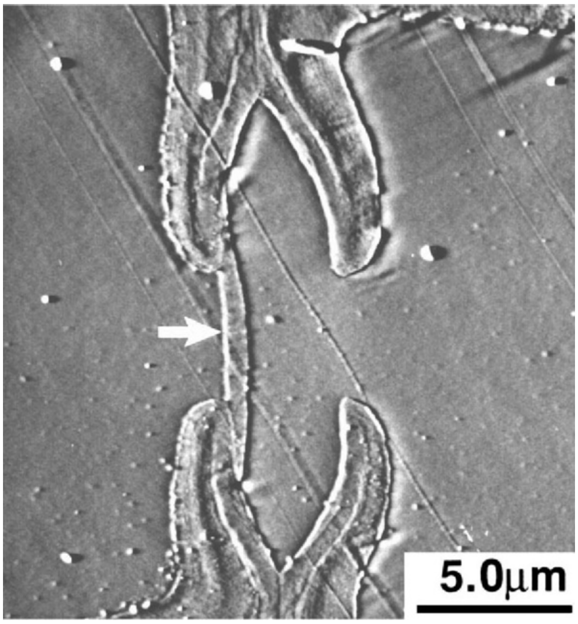


Figure 6
Cross-section of a bordered pit pair. Arrow indicates the aspirated torus blocking one of the pit apertures (after Hanley 1995).

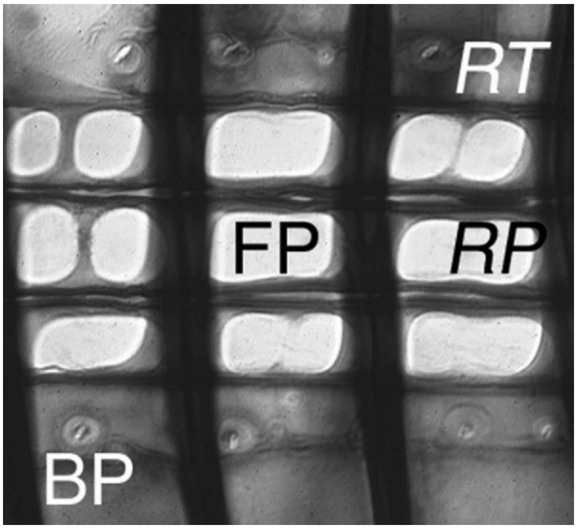


Figure 7
This radial view of a piece of eastern white pine contains both ray parenchyma (*RP*) and ray tracheid (*RT*). FP = fenestriform-type of crossfield pits, and BP = the small bordered pits between the ray tracheids and longitudinal tracheids.

taxodioid, and cupressoid. Figure 7 illustrates a fenestriform crossfield pit in eastern white pine (*P. strobus*). The ray is heterocellular, and the cell types are thereby distinguishable. For more information and examples of crossfield pits the reader is referred to Panshin and de Zeeuw (1980).

(b) *Angiosperm pits*

Since numerous cell types may be present in angiosperms, pitting can be complex, including simple, half-bordered, and full-bordered types. Except for pitting between adjacent vessel elements, the pit appearance on the various types of cells has little diagnostic value. Intervessel pitting (largely on radial walls) is designated as opposite or alternate. Opposite pitting is arranged in more-or-less regular horizontal rows like a checkerboard while alternate pitting is either unordered or arranged somewhat diagonally. A third type of pitting, perforation plates, is found at adjoining ends of vessel elements, having simple openings or more complex structures. In commercial temperate-zone species only two types of perforation plates are found: simple and scalariform. Simple perforation plates are uncomplicated openings surrounded by a thin rim on the ends of the vessel elements, however, scalariform are characterized by numbers of bar-like portions of the cell wall that extend across the vessel openings (Fig. 8). The type of opening and the number of bars are most readily seen on the radial surface and have diagnostic significance.

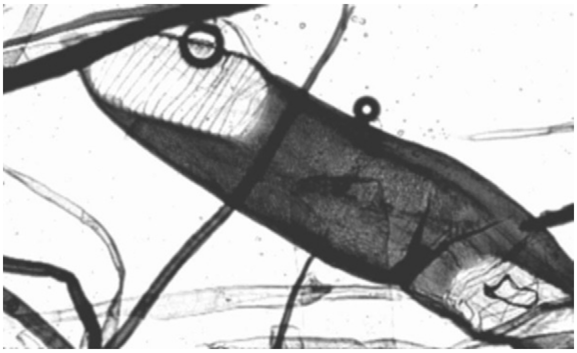


Figure 8
Birch (*Betula* species) vessel element with scalariform perforation plates.

(c) *Cell wall sculpturing*

Wood cell walls may be ornately structured, termed sculpturing. The most frequently observed marking is helical (also called spiral) thickening, which is a coil-like pattern of microfibrillar ridges on the lumen surface. These are a species-specific trait and can be found in gymnosperm tracheids, and in angiosperm vessel elements and vascular tracheids (and occasionally fibers). The feature is distinctive and easily observed using light microscopy. A good example of a North American coniferous species containing this structure is Douglas fir (*Pseudotsuga menziesii*) (Fig. 9).

(d) *Inclusions*

Two types of inclusions are commonly found in lumens of wood cells: mineral crystals and tyloses. Crystals are found in a limited number of species (e.g., iroko (*C. excelsa*), teak (*Tectona grandis*), red mulberry (*Morus rubra*)) and may even be visible to the naked eye (Fig. 10). Tyloses are balloon-like membrane structures that grow into vessel elements from adjacent parenchyma cells (Fig. 11). Tyloses are more common in some species than others; in species where they are usually absent, they may be found in heartwood. Tyloses can block the longitudinal passage of air or other fluids, and may affect the selection of wood for certain uses. Tight cooperage, for example, makes use of white oak (*Quercus species*) with its plentiful tyloses.

3. Reaction Wood

Reaction wood is the generic term used to refer to wood that has atypical anatomical and chemical characteristics, usually from its origin in leaning stems or branches. In gymnosperms this wood is called compression wood and in angiosperms, tension wood. Each form of reaction wood is distinctly different from normal cell tissues.

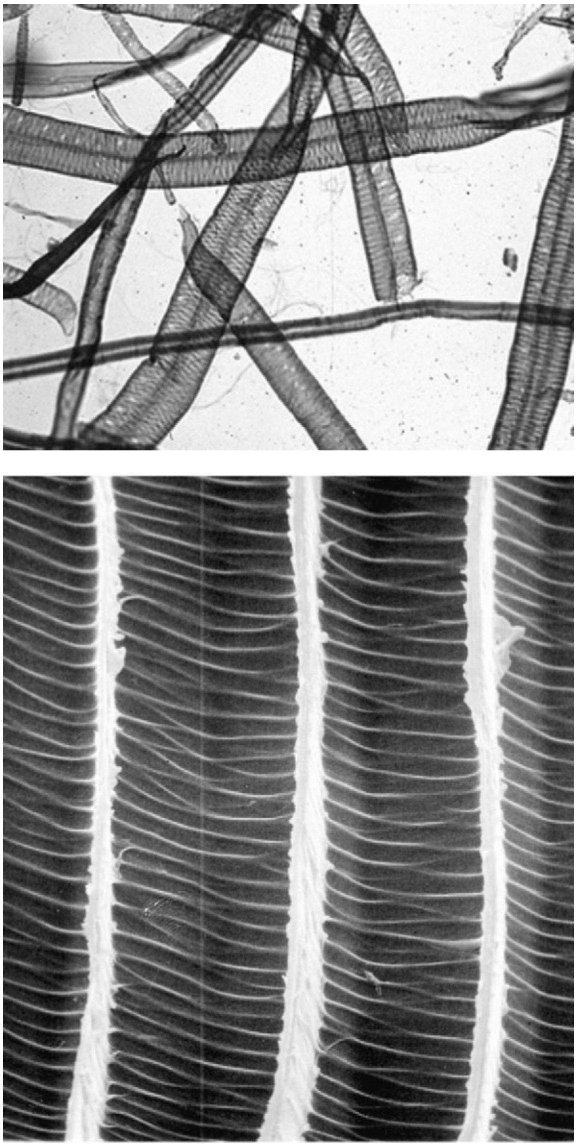


Figure 9
Transmitted light and scanning electron micrographs of Douglas-fir (*P. menzisii*) tracheids showing spiral thickenings.

3.1 Compression Wood

Compression wood is found on the underside of leaning stems and branches of conifers. Its cells are characterized by unusually high lignin and lower cellulose content. The cell walls contain spiral gaps (called "spiral checks") along the steeper S_2 microfibril angle that is a characteristic of these tissues (up to $\sim 45^\circ$ instead of the normal 5–25° from the tracheid axis) (Fig. 12). The cell walls of longitudinal tracheids are

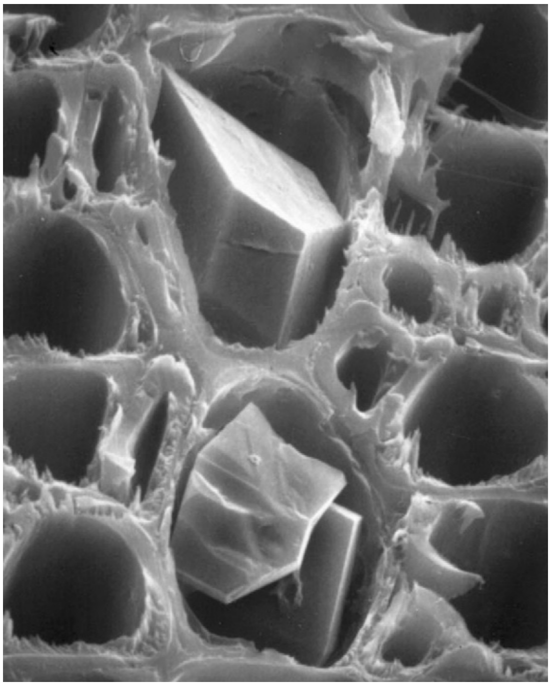


Figure 10
Mineral crystals in iroko (*C. excelsa*) (courtesy of the Institute of Paper Science and Technology).

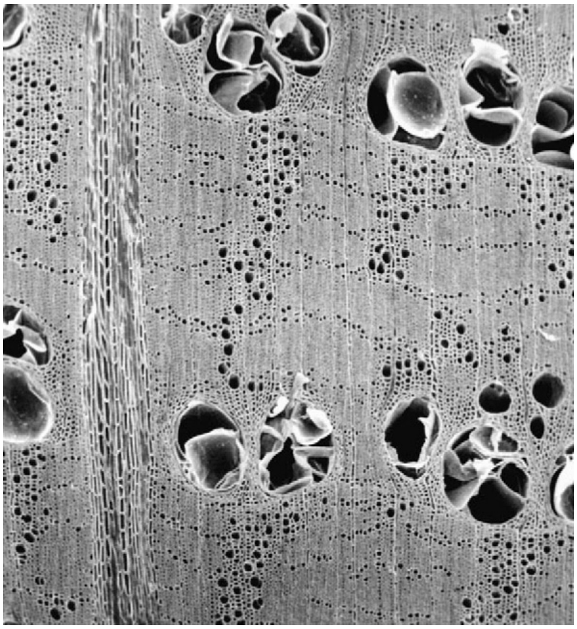


Figure 11
Tyloses in vessel elements of white oak (*Q. alba*) (courtesy of the Institute of Paper Science and Technology).

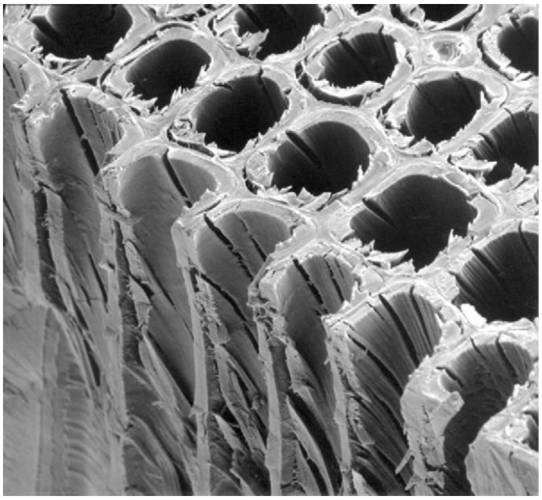


Figure 12
Compression wood in a piece of southern pine showing rounded longitudinal tracheids and spiral checking (micrograph by Kuklinsky R F, Mississippi State University Electron Microscopy Center).

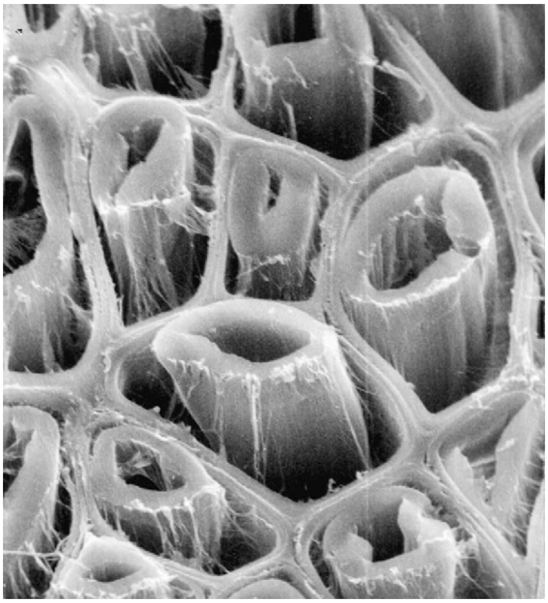


Figure 13
Tension wood showing the G-layer detached from the rest of the cell wall (courtesy of the Institute of Paper Science and Technology).

unusually rounded compared to normal tracheids, and there are intercellular spaces at some cell corners. The cell walls are frequently slightly thicker than normal wood and the S_3 layer is typically lacking.

3.2 Tension Wood

Tension wood is found on the upper side of leaning stems and branches in angiosperms. It contains more cellulose and less lignin, and the fiber structure is markedly different from that of normal wood. Tension wood fibers are characterized by a layer of microfibrils that is almost pure cellulose, the gelatinous or G-layer, and the fibers are known as gelatinous fibers. (Gelatinous fibers may also be found intermixed with normal fibers in oaks even when no reaction wood is present.) The G-layer occurs on the lumen side of the fibers and is easily detached from the rest of the cell wall (Fig. 13), causing a distinctive fuzzy appearance on the surface of machined wood.

Gelatinous fibers can be found in several types of cell wall configurations. The G-layer may occur in addition to the more normal S_1–S_2–S_3 configuration, replace the S_3 layer, or occur in place of both S_2 and S_3. The cellulose microfibrils in the G-layer are oriented at very small angles to the longitudinal fiber axis.

4. Concluding Remarks

The ultrastructure of wood is a very useful diagnostic tool for species identification and also explains some of the gross properties of wood. However, despite the use of many analytical techniques, details of the molecular ultrastructure of wood remain to be resolved.

5. Acknowledgments

The author extends his sincere appreciation to the various individuals and institutions that contributed granted permission to reproduce some of the illustrations in this chapter. These include The Institute of Paper Science and Technology (formerly the Institute of Paper Chemistry), Charles Dunning, Shaune J. Hanley, Derek Gray, and Lennart Salmén. Thanks are also due to the Journal of Pulp and Paper Science, the Mississippi State University Electron Microscopy Center and to Rajai Atalla for reading a portion of an early draft about cellulose structure.

See also: Wood, Constituents of; Wood: Macroscopic Anatomy

Bibliography

Bailey A J 1936 Lignin in Douglas-fir—Composition of the middle lamella. *Ind. Eng. Chem., Anal. Ed.* **8**, 52–5

Baker A A, Helbert W, Sugiyama J, Miles M J 1998 Surface structure of native cellulose microcrystals by AFM. *Appl. Phys. A: Mater. Sci. Process.* **66** (S), S559–63

Browning B L (ed.) 1963 *The Chemistry of Wood.* Interscience, New York, p. 689

Core H A, Côté W A, Day A C 1976 *Wood Structure and Identification,* Syracuse Wood Science Series 6. Syracuse University Press, Syracuse, NY p. 168

Côté W A (ed.) 1965 *Cellular Ultrastructure of Woody Plants.* Syracuse University Press, Syracuse, NY, p. 630

Donaldson L A 1987 S3 lignin concentration in radiata pine tracheids. *Wood Sci. Technol.* **21** (3), 227–34

Dunning C E 1968 An investigation of longleaf pine cell-wall morphology by electron microscopy of single fibers. Ph.D. thesis, Institute of Paper Chemistry, Appleton, WI, p. 409

Faulon J L, Carlson G A, Hatcher P G 1994 A 3-dimensional model for lignocellulose from gymnospermous wood. *Org. Geochem.* **21** (12), 1169–79

Fergus B J, Procter A R, Scott J A N, Goring D A I 1969 The distribution of lignin in sprucewood as determined by ultraviolet microscopy. *Wood Sci. Technol.* **3**, 117–38

Haigler C H, Weimer P J (eds.) 1991 *Biosynthesis and Biodegradation of Cellulose*. Dekker, New York, p. 710

Hanley S J 1995 Application of atomic force microscopy to cellulose, wood, Kraft fibres and paper. Ph.D. thesis, McGill University, Montreal, Canada, p. 129

Hanley S J, Giasson J, Revol J F, Gray D G 1992 Atomic force microscopy of cellulose microfibrils—Comparison with transmission electron microscopy. *Polymer* **33** (21), 4639–42

Hardell H, Westermark U 1981 In: *Ekman-Days Int. Symp. Wood and Pulping Chemistry* Stockholm. **5**, pp. 17–19

Koch P 1972 Utilization of the Southern Pines, Vol. 1, Agriculture Handbook No. 420. US Government Printing Office, Washington, DC p. 734

Koch P 1985 Utilization of Hardwoods Growing on Southern Pine Sites, Vol. 1. Agriculture Handbook No. 605. US Government Printing Office, Washington, DC p. 1418

Kollmann F F P, Côté W A 1968 *Principles of Wood Science and Technology*. Springer, New York p. 592

Meier H 1961 The distribution of polysaccharides in wood fibers. *J. Polym. Sci.* **51**, 11–18

Okuda K, Tsekos I, Brown R M 1994 Cellulose microfibril assembly in *Erythrocladia subintegra* Rosenv—an ideal system for understanding the relationship between synthesizing complexes (TCs) and microfibril crystallization. *Protoplasma* **180** (1–2), 49–58

Page D H, El-Hosseiny F, Bidmade M L, Binet R 1975 Birefringence and the Chemical Composition of Wood Pulp Fibres. Pulp and Paper Reports PPR/137. Pulp and Paper Research Institute of Canada, Pointe Claire, Quebec, Canada

Panshin A J, de Zeeuw C 1980 *Textbook of Wood Technology*, 4th edn. McGraw-Hill, New York, p. 722

Pinto B M (ed.) 1999 *Comprehensive Natural Products Chemistry*. Elsevier, Oxford, Vol. 3

Saka S, Thomas R J, Gratzl J S 1978 Lignin distribution: Determination by energy-dispersive analysis of x-rays. *Tappi* **61** (1), 73–6

Salmén L, Olsson A-M 1998 Interaction between hemicellulosess, lignin and cellulose: Structure-property relationships. *J. Pulp Paper Sci.* **24** (3), 99–103

Sosanwo O A, Fawcett A H, Apperly D 1995 C-13 CP-MAS NMR Spectra of tropical hardwoods. *Polym. Int.* **36** (3), 247–59

Tenkanen M, Tamminen T, Hortling B 1999 Investigation of lignin-carbohydrate complexes in kraft pulps by selective enzymatic treatments. *Appl. Microbiol. Biotechnol.* **51** (2), 241–8

Terashima N, Fukushima K, He L -F, Takabe K 1993 Comprehensive Model of the Lignified Plant Cell Wall. In: Int. Symp. Forage Cell Wall Structure and Digestibility, 1991. American Society of Agronomy-Crop Science, Madison, Wisconsin, pp. 247–70

Timell T E 1967 Recent progress in the chemistry of wood hemicellulosess. *Wood Sci. Technol.* **1**, 45–70

Tsekos I, Okuda K, Brown R M 1996 The formation and development of cellulose-synthesizing linear terminal complexes (TCs) in the plasma membrane of the marine red alga *Erythrocladia subintegra* Rosenv. *Protoplasma* **193** (1–4), 33–45

Weil J, Brewer M, Hendrickson R, Sarikaya A, Ladisch M R 1998 Continuous pH monitoring during pretreatment of yellow-poplar wood sawdust by pressure cooking in water. *Appl. Biochem. Biotechnol.* **70** (2), 99–111

Westermark U, Lidbrandt O, Eriksson I 1988 Lignin distribution in spruce (*P. abies*) determined by mercurization with SEM-EDXA technique. *Wood Sci. Technol.* **22**, 243–50

Wood J R, Goring D A I 1971 The distribution of lignin in stem wood and branch wood of Douglas fir. Technical Paper T95. *Pulp Pap. Mag. Can.* **72** (3), 61–8

T. E. Conners
Forest Products Laboratory, Mississippi, USA

This Page Intentionally Left Blank

LIST OF CONTRIBUTORS

Contributors are listed in alphabetical order together with their addresses. Titles of articles that they have authored follow in alphabetical order. Where articles are coauthored, this has been indicated by an asterisk preceding the title.

Aaronson, H. I.
Carnegie Mellon University
Pittsburgh, Pennsylvania
USA and
Monash University
Clayton, Victoria
Australia
**Ferrous Alloys: Overview*

Ahmad, Z.
Quaidi-i-Azam University
Islamabad
Pakistan
Ceramic-modified High-temperature Polymers

Ashby, M. F.
University of Cambridge
UK
Composite Materials, Microstructural Design of

Ballarini, R.
Case Western Reserve University
Cleveland, Ohio
USA
**Shell: Properties*

Baltá-Calleja, F. J.
Instituto de Estructura de la Materia
CSIC
Madrid
Spain
**Polymer Crystallization: General Concepts of Theory and Experiments*

Bassett, D. C.
University of Reading
UK
Polymer Spherulites

Beaudoin, A. J.
University of Illinois
Urbana, Illinois
USA
**Deformation Processing: Texture Evolution*

Blair, M.
SFSA
Barrington, Illinois
USA
Stainless Steels: Cast

Booth, C.
University of Manchester
UK
Cyclic Polymers, Crystallinity of

Bousfield, D. W.
University of Maine
Orono, Maine
USA
Paper: Porosity

Brooker, L. R.
Murdoch University
WA
Australia
**Marine Teeth (and Mammal Teeth)*

Caron, R. N.
Olin Corporation
New Haven, Connecticut
USA
Copper Alloys: Alloy and Temper Designation

Cawley, J. D.
Case Western Reserve University
Cleveland, Ohio
USA
Binary Oxide Ceramics: Al_2O_3, ZrO_2, Structure and Properties of

Chandrasekhar, S.
Centre for Liquid Crystal Research
Bangalore
India
Discotic Liquid Crystals: Overview

Chen, I.-W
University of Pennsylvania
Philadelphia, Pennsylvania
USA
**Si-AlON Ceramics, Structure and Properties of*

Conners, T. E.
Forest Products Laboratory
Mississippi
USA
Wood: Ultrastructure

Crook, P.
Haynes International Inc.
Kokomo, Indiana
USA
**Cobalt Alloys: Alloying and Thermomechanical Processing*

Das, S.
University of Michigan
USA
**Tissue Engineering Scaffolds*

Demus, D.
Halle
Germany
Calamitic Liquid Crystals

Donaldson, L. A.
Forest Research
Rotorua
New Zealand
**Wood, Constituents of*

Effenberg, G.
Materials Science International Services GmbH
Stuttgart
Germany
**Phase Diagrams: Data Compilation*

Engler, O.
VAW Aluminium
Bonn
Germany
**Deformation Processing: Texture Evolution*

Ezquerra, T. A.
Instituto de Estructura de la Materia
CSIC
Madrid
Spain
**Polymer Crystallization: General Concepts of Theory and Experiments*

Feiring, A. E.
DuPont
Wilmington, Delaware
USA
Fluorine-containing Polymers

Feltz, A.
Deutschlandsberg
Austria
Oxide Glasses

Finney, J. L.
University College London
UK
Ice: Structures

Froes, F. H.
University of Idaho
Moscow, Idaho
USA
Titanium: Alloying
Titanium Alloys: Alloy Designation System

Garrison, Jr., W. M.
Carnegie Mellon University
Pittsburgh, Pennsylvania
USA
**Ferrous Alloys: Overview*
Steels: Classifications

Gido, S. P.
University of Massachusetts
Amherst, Massachusetts
USA
Block Copolymers, Structural Characterization of

Giesa, R.
University of Bayreuth
Germany
**High-temperature Stable Polymers*

Goodall, B. L.
Albemarle Corporation
Baton Rouge, Louisiana
USA
Cycloaliphatic Polymers

Gray, G. W.
Wimborne
UK
Liquid Crystals: Overview

Green, D. J.
Pennsylvania State University
University Park, Pennsylvania
USA
Porous Ceramics including Fibrous Insulation, Structure and Properties of

Hampshire, S.
University of Limerick
Ireland
Oxynitride Glasses

Harmer, M. P.
Lehigh University
Bethlehem, Pennsylvania
USA
**Nanoscale Ceramic Composites*

Hedrick, J. L.
IBM Almaden Research Center
San Jose, California
USA
**Hybrid Dendrimer Star-like Polymers*

Heppke, G.
Technische Universität Berlin
Germany
**Nematic Discotic Liquid Crystals*

Heuer, A. H.
Case Western Reserve University
Cleveland, Ohio
USA
**Shell: Properties*

Hoffmann, M. J.
Universität Karlsruhe
Germany
Si_3N_4 Ceramics, Structure and Properties of

Hogan, L. M.
University of Queensland
Brisbane, Queensland
Australia
Crystals, Dendritic Solidification of

Hollister, S. J.
University of Michigan
USA
**Tissue Engineering Scaffolds*

Hudd, R. C.
Porthcawl
UK
Sheet Steel: Low Carbon

Ito, H.
Tokyo Institute of Technology
Japan
Injection Molded Semicrystalline Polymers: Structure Development and Modeling

Jensen, D. J.
Risø National Laboratory
Roskilde
Denmark
Annealing Textures

Jonas, U.
Max-Planck-Institut für Polymerforschung
Postfach, Mainz
Germany
**Polymeric Discotic Liquid Crystals*

Jones, R. L.
Pennsylvania State University
University Park, Pennsylvania
USA
**Polymer Surfaces: Structure*

Kamat, S.
Case Western Reserve University
Cleveland, Ohio
USA
**Shell: Properties*

Klarstrom, D.
Haynes International Inc.
Kokomo, Indiana
USA
**Cobalt Alloys: Alloying and Thermomechanical Processing*

Klein, L. C.
Rutgers University
Piscataway, New Jersey
USA
**Polymer–Ceramic Nanocomposites: Polymer Overview*

Kleman, M.
Université Pierre-et-Marie-Curie
Paris
France
Cholesteric Liquid Crystals: Defects

Krüerke, D.
Technische Universität Berlin
Germany
**Nematic Discotic Liquid Crystals*

Kuhn, L. T.
Children's Hospital and Harvard Medical School
Boston, Massachusetts
USA
Bone Mineralization

Kumar, S. K.
Pennsylvania State University
University Park, Pennsylvania
USA
**Polymer Surfaces: Structure*

Lee, A. P.
Murdoch University
WA
Australia
**Marine Teeth (and Mammal Teeth)*

Lubensky, T. C.
University of Pennsylvania
Philadelphia, Pennsylvania
USA
Chiral Smectic Liquid Crystals

Macey, D. J.
Murdoch University
WA
Australia
**Marine Teeth (and Mammal Teeth)*

Margadonna, S.
University of Sussex
Brighton
UK
**Fullerenes*

Marsh, H.
North Shields
UK
Carbon Mesophase

Massalski, T. B.
Carnegie Mellon University
Pittsburgh, Pennsylvania
USA
**Phase Diagrams*

McDonald, A. G.
Forest Research
Rotorua
New Zealand
**Wood, Constituents of*

Mitchell, G. R.
University of Reading
UK
Structure of Polymer Glasses: Short-range Order

Militzer, M.
University of British Columbia
Vancouver, British Columbia
Canada
Austenite Decomposition: Overall Kinetics during Isothermal, and Continuous Cooling Transformation

Muddle, B. C.
Monash University
Clayton, Victoria
Australia
**Martensite*

Nie, J. F.
Monash University
Clayton, Victoria
Australia
**Martensite*

Novotny, P. M.
Carpenter Technology Corporation
Reading, Massachusetts
USA
Tool and Die Steels

Padilha, A. F.
Universidade Federal Fluminense
Rio de Janeiro
Brazil
**Precipitation from Austenite*

Perkins, R.
Syracuse University
New York
USA
Paper: Structure

Pinckney, L. R.
Corning Incorporated
Corning, New York
USA
Glass Ceramics

Prassides, K.
University of Sussex
Brighton
UK
**Fullerenes*

Reynolds, Jr., W. T.
Virginia Polytechnic Institute and State University
Blacksburg, Virginia
USA
Bainite

Rios, P. R.
Universidade Federal Fluminense
Rio de Janeiro
Brazil
**Precipitation from Austenite*

Rundman, K. B.
Michigan Technological University
Houghton, Michigan
USA
Cast Irons

Schmidt, H.-W.
University of Bayreuth
Germany
**High-temperature Stable Polymers*

Shiflet, G. J.
University of Virginia
Charlottesville, Virginia
USA
Pearlite

Shuba, R.
University of Pennsylvania
Philadelphia, Pennsylvania
USA
**Si-AlON Ceramics, Structure and Properties of*

Solomon, H. D.
General Electric R & D Center
Schenectady, New York
USA
Stainless Steels: Duplex

Spiess, H. W.
Max-Planck-Institut für Polymerforschung
Postfach, Mainz
Germany
**Polymeric Discotic Liquid Crystals*

Starke, Jr., E. A.
University of Virginia
Charlottesville, Virginia
USA
Aluminum: Alloying
Aluminum Alloys: Alloy, Heat Treatment, and Temper Designation

Steurer, W.
Federal Institute of Technology
Zurich
Switzerland
Crystal Structures of the Elements

Su, X.
Case Western Reserve University
Cleveland, Ohio
USA
**Shell: Properties*

Subramoney, S.
DuPont Company
Wilmington, Delaware
USA
Carbon Nanotubes

Tanaka, K.
Hokkaido University
Sapporo
Japan
Chalcogenide Glasses

Thoma, D. L.
Los Alamos National Laboratory
New Mexico
USA
Intermetallics: Laves Phases

Thompson, G. S.
Lehigh University
Bethlehem, Pennsylvania
USA
**Nanoscale Ceramic Composites*

Trollsås, M.
Candescent Technologies
San Jose, California
USA
**Hybrid Dendrimer Star-like Polymers*

Usuki, A.
Toyota Central R&D Labs
Aichi
Japan
Clay-based Polymer Nanocomposites

Varin, R. A.
University of Waterloo
Ontario
Canada
Intermetallics: Crystal Structures

Vincent, J. F. V.
University of Reading
UK
Cuticle

Webb, J.
Murdoch University
WA
Australia
**Marine Teeth (and Mammal Teeth)*

Weber, J. H.
Special Metals Corporation
Huntington, West Virginia
USA
Nickel Alloys: Nomenclature
Nickel-based Superalloys: Alloying

Weiner, S.
Weizmann Institute of Science
Rehovoth
Israel
**Teeth: Structure/Mechanical Design Strategies*

Westengen, H.
Norsk-Hydro ISI
Porsvrunn
Norway
Magnesium: Alloying

Wheeler, E.
North Carolina State University
Raleigh, North Carolina
USA
Wood: Macroscopic Anatomy

Whitmore, M. D.
Memorial University of Newfoundland
St. John's Newfoundland
Canada
Block Copolymer Phase Behavior

Wilkinson, A. J.
University of Oxford
UK
Deformation Textures

Windle, A. H.
University of Cambridge
UK
Liquid Crystalline Polymers: An Introduction

Wojcik, A. B.
Hybrid Glass Technologies
Princeton, New Jersey
USA
**Polymer–Ceramic Nanocomposites: Polymer Overview*

Wright, J. D.
University of Kent
Canterbury
UK
Crystal Engineering

Young, J. F.
University of Illinois
Urbana, Illinois
USA
Portland Cements

Zaslansky, P.
Weizmann Institute of Science
Rehovoth
Israel
**Teeth: Structure/Mechanical Design Strategies*